AD_____

Award Number: W81XWH-04-1-0571

TITLE: Detecting and Targeting Oncogenic Myc in Breast Cancer

PRINCIPAL INVESTIGATOR: Linda Z. Penn, Ph.D.

CONTRACTING ORGANIZATION: University Health Network
Toronto, Ontario M5G 2M9

REPORT DATE: June 2007

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01/06/07	Final	10 May 2004 – 9 May 2007

4. TITLE AND SUBTITLE

Detecting and Targeting Oncogenic Myc in Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0571

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Linda Z. Penn, Ph.D.

E-Mail: lpenn@uhnres.utoronto.ca

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University Health Network
Toronto, Ontario M5G 2M9

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT: Deregulation of the cellular myc proto-oncogene is one of the strongest activators of tumorigenesis and understanding the target genes regulated by this transcription factor in cancer etiology will clearly mark a key advance. Here we identify the non-coding RNA H19 as a Myc-induced gene that plays a functional role in breast cancer development. We have also developed antibody reagents to TRRAP, a cofactor that collaborates with Myc to drive tumorigenesis. To determine which target genes are co-regulated by Myc and TRRAP in breast cancer, we have optimized a ChIP-on-chip protocol (Chromatin immunoprecipitation coupled with microarray technology) that can be used with breast cancer cells and primary specimens to achieve both sensitive and specific results. We have also mapped two regions of TRRAP that interact with Myc as shown by co-immunoprecipitation. Our work will impact breast cancer research in several ways. The refined ChIP-on-chip protocol will be used by ourselves and others to conduct breast cancer research focused on specific transcription factors, such as Myc, as well as other regulatory mechanisms at the level of chromatin (e.g. replication, DNA repair). In addition to several valuable antibody reagents, we have identified a novel Myc binding domain of TRRAP and we have identified H19 as a molecular target for the development of novel anti-cancer therapeutics.

15. SUBJECT TERMS
Myc oncogene, Transcription factor, Chromatin immunoprecipitation, microarray, node-negative breast tumor, inhibitor screen, gene expression

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU	148	USAMRMC

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Introduction

Background: Deregulation of the cellular myc proto-oncogene is one of the strongest activators of tumorigenesis and understanding the target genes and pathways regulated by this transcription factor in cancer etiology will clearly mark a key advance. Myc expression and activity are highly restricted in normal human mammary epithelial cells (HMECs), but unleashed and deregulated in cells of malignant transformations. Because the protein product is identical in normal and tumor cells, three major issues arise. First, a definitive approach to detect oncogenic Myc in primary tumor specimens is severely lacking and long overdue. This issue has plagued the field during the two decades since Myc was first discovered. Second, it remains unclear whether Myc *function* is different in normal and tumor cells. Myc may regulate the same subset of target genes in both settings, but in a more robust manner in tumor cells. By contrast, deregulated, overexpressed Myc protein may bind and regulate an additional unique set of target genes in tumor cells. This issue has not yet been explored. Third, although it is clear that inhibiting Myc can trigger tumor regression and eradication in animal models, few initiatives are underway to target Myc as a therapeutic approach for human disease. Because Myc protein in normal and tumor cells is indistinguishable, it is thought that anti-Myc inhibitors would have little to no tumor specificity or therapeutic index. Clearly a novel approach is required.

Hypothesis: Our 'idea' is that in addition to the target genes regulated by Myc in non-transformed cells, constitutively activated and overexpressed Myc protein in tumor cells will directly bind and regulate a unique set of target genes that directly contribute to the carcinogenic process. For example, at high levels of expression, Myc may bind low affinity sites and regulate a distinct cohort of targets by a unique mechanism of action. By identifying this transformation specific subset of Myc target genes we aim to develop a diagnostic tool to identify oncogenic Myc *activity* in breast tumor cells. We also aim to develop a unique anti-cancer therapeutic that will potentially target this unique transforming activity of Myc. The TRRAP cofactor has been shown to be essential for Myc to drive transformation. This suggests blocking Myc:TRRAP interaction will inhibit the carcinogenic program directed by oncogenic Myc. By conducting the experiments outlined in this proposal we will test a unique hypothesis and will make significant contributions to the molecular diagnosis and treatment of breast cancer that can be applied to the clinic in a timely manner

Specific Aims:

1) Identify tumor-specific, directly-regulated Myc target genes in transformed HMECs and develop a definitive diagnostic tool to detect oncogenic Myc activity in breast cancer.
2) Isolate small molecular weight inhibitors that can disrupt Myc:TRRAP interaction in vivo and identify Myc:TRRAP co-bound target genes in breast cancer.

Body

With support from the DOD, the research outlined in the original proposal has progressed in a steady and productive manner throughout the course of this grant despite several unexpected technical problems along the way. To delineate the accomplishments to date, the tasks outlined in the original Statement of Work of the proposal are itemized below (italics) and a progress report for each task provided.

Statement of Work

Task 1: Identify tumor-specific, directly-regulated Myc target genes in transformed HMECs (months 1-36)

a. *Directly compare fresh and formalin-fixed tissue for efficacy of Chip-on-chip assay using samples, including archived samples, prepared by the Andrulis group (months 1-6)*

Completed in year 1: We anticipated that archived formalin-fixed tissue could be used for ChIP-on-chip analysis, however, over the course of the first year, we learned that formalin fixation conducted in pathology labs is not similar to the formaldehyde fixation used in our research lab as an early step of the ChIP-on-chip procedure. We suspect the issue has more to do with the heterogeneity in the methodology used by different technicians/residents over the years in various pathology labs (formalin formulation, time of fixation, etc). To overcome this problem, we evaluated whether frozen tissue could be used as an alternate source. We are pleased to report that we can achieve robust ChIP-on-chip results using 0.03g of fresh, frozen tissue. Importantly, we can move forward with the study, using frozen instead of formalin-fixed primary tissue for our ChIP-on-chip analysis.

b. *Evaluate minimum numbers of cells that is required for quality assured Chip-on-chip of tissue sections, test reproducibility (months 6-12)*

Completed in years 1 to 3: When we first launched this research we had several parameters of the ChIP-on-chip assay to optimize to ensure we could achieve the high sensitivity and specificity required to apply this novel technology to breast cancer cell lines and primary patient specimens. In year 1 we compared and modified two amplification methods to determine how few cells are required to achieve efficient ChIP-on-chip results without introducing extreme bias due to amplification. In year 2, several additional parameters were evaluated and conditions determined to ensure outstanding reproducibility of this new technology. This data was well-received when it was presented in 2006 at the AACR Annual Meeting and a Cold Spring Harbor Systems Biology Meeting.

In year 3, our goal was to apply our optimized ChIP-on-chip assay to map the genomic localization of Myc-binding using frozen breast tissue. However, the archived tissue from the ANN tissue bank of our collaborator (Andrulis) showed enormous variability, that we assume was due to the lengthy storage times of these samples. These variability issues were not evident in fresh, frozen tissue used as a positive control. An alternative source of ANN was not readily available, so we decided to evaluate our optimized ChIP-on-chip protocol on breast cancer cell lines. However we were not able to achieve the sensitivity and specificity required. Even by increasing the number of these adherent breast cells to very high levels (60 million cells per ChIP reaction), the signal to noise ratio was not sufficiently robust to move forward from ChIP to the ChIP-on-chip experiments. We have overcome this issue very recently by changing from a probe sonicator approach to the far more efficient Bioruptor sonication system. We were only able to purchase the Bioruptor Instrument about a month ago after applying and securing funding from our Institutional Capital Equipment Competition at the beginning of 2007. Because of this delay, we moved the experiments forward and completed our study to test our optimized parameters for ChIP-on-chip using non-adherent HL60 cells. These cells have high levels of endogenous Myc expression and

can be grown in bulk as suspension cells, thus ChIP-on-chip can be conducted despite the poor recovery of ChIP material using the old probe sonicator method. We have completed the experiments associated with this work and have recently submitted a manuscript, entitled, "Optimization of experimental design parameters for high-throughput chromatin immunoprecipitation studies" that is presently under review for publication.

c. *Conduct the Myc specific Chip-on-chip assay in sixplicate with reverse dye labeling on one sample that has been processed for lasar capture microdissection for normal and tumor isogenic matched sample, in duplicate (months 12 –18)* Yet to do*
d. *Conduct Myc specific Chip-on-chip assay on 2 additional genetically similar patient samples (months 12-18)* Yet to do*
e. *Conduct Myc specific Chip-on-chip on 3 similar samples whose underlying genetic abnormalities are distinct from the first series of 3 (months 18-24)* Yet to do*
f. *Conduct data analysis to evaluate how target genes compare between isogenic normal and tumor matched material and between groups that harbor distinct genetic abnormalities and between all samples to identify a cohort commonly bound and regulated by Myc in tumor but not normal tissue. (months 24 –32)* Yet to do*
g. *Further evaluate the diagnostic potential of this common cohort and evaluate whether cDNA expression profiling shows these targets are similarly regulated in ANN tumors of a specific subtype or genetic background (months 24 –32)* Yet to do*
h. *Extend analysis to tissue arrays using in situ hybridization or IHC (months 24-36)* Yet to do*
i. *Evaluate cDNA expression array data for genes identified as regulated by Myc in the MCF10A system (months 12-36)* Yet to do*

*Due to unexpected difficulties of working with archived formalin, and then frozen breast specimens, along with the added complexity of breast cells and the need for the Bioruptor system, we were not able to conduct the ChIP-on-chip experiments proposed. However, we are dedicated to the breast cancer question and these optimized methods will now be applied to breast cancer. Indeed, recent ChIP results show that with the Bioruptor we can achieve an excellent signal to noise ratio with only 10 million cells. A further reduction in cell number is likely achievable and is presently under evaluation. With these recent technical advances, we are now in a strong position to conduct Myc-specific ChIP-on-chip on both breast cancer cell lines and primary fresh, frozen biopsies. We have recently established collaborations with Dr. Susan Done who has established a specific breast tumour bank at our Institute in recent years, that specializes in the preservation and analysis of small primary breast tumours (<2cm) as well as charting patient treatment and outcome. We look forward to working with Susan and her new state-of-the-art facility. We will analyze these tumours for elevated Myc expression (using affymetrix expression arrays) and then further evaluate whether elevated expression is due to amplification of the c-myc locus using the Taqman and FISH assays that we have developed for this purpose. As proposed we will then conduct the ChIP-on-chip experiments, now that we have the protocol optimized for this purpose. It is because of this DOD funding that we could move this technology forward for the analysis of patient specimens. These advances will enable our own research in this area, but will also be used by others whose goals are to use ChIP-on-chip in their breast cancer research. We will acknowledge the support of the U.S. DOD when these studies are published.

Task 2: Establish MCF10A cell system and identify tumor-specific, directly-regulated Myc target genes (months 1-36)

a. *Introduce ectopic Myc expression in the MCF10A cells and evaluate biological effect at the level of proliferation, apoptosis induction in standard culture conditions (months 1-6)* Completed in year 1. We have ectopically expressed Myc in the MCF10A cells and have characterized cell growth and death, as proposed. Myc potentiates both cell proliferation and apoptosis. This was shown in our year one report.

b. *Conduct Myc ChIP-on-chip analysis and cDNA array analysis to identify target genes directly bound and regulated by Myc under asynchronous conditions (months 1-12)* Completed in years 1 and 2. We conducted in silico analysis on the expression array data from the ANN samples of Irene Andrulis and showed that H19 expression was highly associated with elevated Myc expression in these tumours. We conducted ChIP analysis in breast cancer cell lines and identified H19 as a novel Myc target gene regulated by Myc in MCF10As. We were able to conduct ChIP using Myc-specific antibody while assaying a specific target gene, such as H19, but the method was not robust enough for ChIP-on-chip (see above). With this positive control established, we went on to study the regulation of H19 by Myc and to study the role of H19 in breast cancer. We further showed that Myc induction of H19 plays a role in Myc induced transformation by conducting siRNA knock-downs in breast cancer cell lines that do and do not express H19. Indeed, the loss of H19 inhibits full transformation potential of these breast cancer cells. Thus we show that we have identified at least one novel Myc-induced gene (H19), using the breast cancer cell system, that plays an important role in transformation. Only a handful of Myc target genes have been shown to play a role in transformation, thus it was important to focus on this single important target and ensure the publication of this information in a timely manner. To this end, a manuscript describing this work was been published in *Cancer Research* (Appendix 1). In addition, this manuscript was highlighted in *Nature Reviews Cancer* in the July 2006 issue (Appendix 2). Using the MCF10A system and our ChIP-on-chip analysis, we aim to identify similar Myc targets that are, like H19, Myc-regulated and key to breast cancer development, as originally proposed. We and others are now targeting H19 for the development of anti-breast cancer therapeutics.

c. *Into the control and Myc expressing cells introduce activatable erbB2 and p53-DD (months 6-12)*. No longer necessary. Originally it remained unclear whether deregulated Myc alone would be able to transform the MCF10A cells, as assessed by anchorage-independent growth in soft agar. It was anticipated that additional genetic lesions, such as erbB2 and/or p53-DD would be necessary and was proposed. However, these additional genetic lesions are not required. Ectopic Myc alone is able to transform the MCF10A cells. Thus our experimental model system was established in the absence of erbB2 and p53-DD.

d. *Assay cells for growth in soft agar as well as proliferation and apoptosis assays (months 12-18)* Complete. The MCF10A +/- Myc cells have been assayed thoroughly and form the basis of our future work. Myc potentiates proliferation, apoptosis and growth in soft agar of these cells. Several strains of MCF10A were evaluated and all behaved similarly with respect to Myc effects on cellular growth, death and tumorigenesis. This is an enormous break-through as we can measure the molecular

mechanism of Myc-induced transformation as ectopic Myc expression converts these immortal, non-transformed cells to a further transformed state as measured by anchorage independent growth. This system will be used to identify the critical transformation-associated genes regulated by Myc.

e. *Assay Myc ChIP-on-chip and cDNA arrays on these cells expressing ectopic Myc and/or erbB2 and/or p53-DD, under asynchronously growing conditions (months 18-32)* Yet to do#.
f. *Assay more transformed cells for invasion, polarity, morphology properties (months 18 – 32)* Yet to do#.
g. *Test for genetic abnormalities and their effects on cell differentiation in matrigel (months 24-36)* Yet to do#.
h. *Assay all cells for Myc Chip-on-chip and cDNA arrays when grown in soft agar (months 24-36)* Yet to do#.

#We conducted the ChIP-on-chip assays with the MCF10A cells, with and without ectopic Myc expression, but it was with this work that we further realized the sensitivity and specificity of the assay was insufficient. After conducting 10 biological replicates we could see that we had achieved signal over noise, as evident by volcano plots, but the signal did not achieve statistical significance, despite the high number of replicates. Rather than continuing to repeat the assay, we decided we had to trouble-shoot the protocol once again to determine the problem and then to overcome this rate-limiting step. As discussed above, this turned out to be the recovery of specific cross-linked material from the breast cancer cells and this was solved by moving to a more efficient sonication system (Bioruptor). Thus, we have a robust cell system and ChIP-on-chip protocol that are working well. We have been able to develop the assay to work for both Myc and TRRAP-specific antibodies (see Figure 1). We will acknowledge the DOD for funding support when this work is complete and published.

Task 3:Isolate small molecular weight inhibitors that can disrupt Myc:TRRAP interaction in vivo and identify Myc:TRRAP co-bound target genes in breast cancer.

a. *Develop and test antibodies to TRRAP for ChIP (months 1-12)* Completed. After much trouble-shooting we are able to readily ChIP TRRAP from live cells using a homemade antibody. The specificity of the antibody has been rigorously evaluated and the ChIP results are robust and quantitative. This has enable TRRAP ChIP-on-chip to be conducted in the MCF10A cell system described above. We aim to also advance these assays to primary patient material, as proposed. By this approach the target genes that are co-bound by Myc and TRRAP, that are essential for breast cancer transformation, will be identified.
b. *Develop Myc and TRRAP interacting fragments in new screening system and evaluate interaction and susceptibility to inhibition with TRRAP polypeptide (months 1-12)* Conducted in Year 2 and 3. Before setting up the screening system we needed to ensure the fragment of TRRAP shown to interact with Myc did indeed interact with Myc in vivo. This has been evaluated and is submitted for publication (Appendix 3). Targeting Myc as a novel anti-cancer therapeutic has enormous potential (for more information, see our review Appendix 4). With these results we went forward to establish the screening system, as proposed, but the system did not work with our positive and negative controls (See Figure 2A, F8 and F1 respectively). The RTA was modified,

however after conducting the screen we characterized the few positive hits and revealed that they were all false positives. This was disappointing and frustrating.

c. *Conduct screen and test positives in the RTA against Myc:TRRAP as well as other interactors of Myc and other interactors of TRRAP to evaluate specificity of the inhibitor (months 12-18)* Conducted in year 3. The hits were shown to be false positives.

d. *Advance inhibitors to mammalian cell assays, including growth, transformation DNA binding using a Chip-on-chip approach (months 18-36)* Not achieved^

^With the disappointing results of our screen, and in the interest of time, we decided it was not worthwhile to further pursue the RTA as a screening system. The rate of false positives was too high and there was insufficient time to re-tool the assay and conduct the screen again before the grant expired. We decided to be productive in the time remaining (4 months) we would work towards identifying another region of Myc:TRRAP interaction that could potentially also be targeted for anti-cancer therapeutics. Our friend and collaborator Michael Cole shared information with us that he had identified that amino acids 20 to 40 of Myc can also bind to TRRAP (for details, see his recent review article[1]), so we evaluated the region of TRRAP that might be involved. This builds on our previous work and enabled us to further exploit reagents we had already made in the lab. Indeed, we have been able to identify another region of TRRAP, Myc Binding Domain 2 (MBD2), that interacts with Myc (see Figure 2B).

Key Research Accomplishments

- We have exploited the ANN expression data and identified H19 as a bone fide Myc target gene that contributes to transformation of breast cancer. This has been published.
- We have established that deregulated Myc expression is sufficient to transform MCF10A cells as defined by anchorage independent growth in soft agar. This system is ideal to further identify the entire subset of genes Myc regulates to transform human mammary epithelial cells.
- We have optimized the ChIP-on-chip protocol and have been able to successfully trouble-shoot several issues associated with ChIP-on-chip in breast cell and primary patient tissue.
- Myc and TRRAP interaction domains have been thoroughly analyzed and fully characterized. This has resulted in a publication that has been submitted.
- We have identified a novel region of TRRAP (MBD2) that interacts with Myc as shown by ectopic expression followed by co-immunoprecipitation experiments.

Reportable Outcomes

- Poster presented at the Era of Hope Department of Defense Breast Cancer Research Program Meeting, June 8 – 11, 2005, Pensylvania Convention Centre, Philadelphia, Pennsylvania
- Barsyte-Lovejoy D, Lau SK, Boutros PC, Khosravi F, Jurisica I, Andrulis IL, Tsao MS, Penn LZ. The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. Cancer Res. 2006 May 15;66(10):5330-7. (see Appendix 1)
- Nature Reviews Cancer – The Silent Messenger. July (6) 2006:488. Highlight of Cancer Research article above (see Appendix 2)
- Ponzielli R, Katz S, Barsyte-Lovejoy D, Penn LZ. Cancer therapeutics: targeting the dark side of Myc. Eur J Cancer. 2005 Nov;41(16):2485-501. Review. (see Appendix 3)
- Lilia Kaustov, Sigal Katz, Cynthia S.W. Ho, Romina Ponzielli, Shili Duan, Steve McMahon, Michael D.Cole, Linda Z. Penn, and Cheryl H. Arrowsmith , Characterizing cMyc MBII and its interaction with TRRAP, submitted and under review, (see Appendix 4)

•Optimizing ChIP-on-chip and reporting on key parameters of technology – presented at AACR Annual Meeting, Washington, DC, April 1-5, 2006 (Penn oral presentation, Boutros poster presentation); also presented at Cold Spring Harbor Labs, System Biology Conference, Cold Spring Harbor, NY, Mar 23 –26, 2006 (Boutros poster presentation)
•Oral presentation by Penn at "Reasons for Hope" Conference, sponsored by the Canadian Breast Cancer Research Alliance, Montreal, Quebec, May 6-8, 2006
•Submission of manuscript entitled, "Optimization of experimental design parameters for high-throughput chromatin immunoprecipitation studies". (see Appendix 5)

Conclusions

Several publications have arisen from this research and several additional significant contributions are expected.

1. We established a useful model system to recapitulate the transition from non-transformed (MCF10A cells) to transformed (MCF10A cells with deregulated Myc expression) human mammary epithelial cells. This system is the best we could have hoped for. It is simple and does not involve the need to introduce additional genetic abnormalities to achieve anchorage independent growth. This enables the genetics to be readily interpreted. Indeed, this system enabled the Myc target gene H19 to be quickly evaluated for it mechanism of regulation and importantly, its role in breast cancer transformation.
2. We have developed the technologies to determine which tumors have high Myc expression due to gene amplification (Taqman, FISH) and to determine the genomic localization of Myc using a significantly optimized ChIP-on-chip protocol for primary and established cells derived from breast cancers.
3. We have developed critical antibodies that recognize TRRAP and work in ChIP to identify TRRAP-bound target genes in a sensitive and specific manner. These valuable reagents will be used for further ChIP-on-chip analyses, as proposed.
4. We have established two regions of Myc and TRRAP interaction.

Salaries paid from this grant: Sigal Katz, Paul Boutros, Fereshteh Khosravi, Angelina Stojanova. In addition, Romina Ponzielli and Dalia Barsyte-Lovejoy worked on aspects of this work, but they had secured their own scholarships to cover their salaries.

References

1. Cowling, V. H. and Cole, M. D. Mechanisms of transcriptional activation by the Myc oncoproteins. Semin Cancer biol 16, 242-52 (2006).

Appendices

Appendix 1: Barsyte-Lovejoy D, Lau SK, Boutros PC, Khosravi F, Jurisica I, Andrulis IL, Tsao MS, Penn LZ. The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. Cancer Res. 2006 May 15;66(10):5330-7.

Appendix 2: Nature Reviews Cancer – The Silent Messenger. July (6) 2006: 488. Highlight of Cancer Research article above.

Appendix 3: Ponzielli R, Katz S, Barsyte-Lovejoy D, Penn LZ. Cancer therapeutics: targeting the dark side of Myc. Eur J Cancer. 2005 Nov;41(16):2485-501. Review.

Appendix 4: Lilia Kaustov, Sigal Katz, Cynthia S.W. Ho, Romina Ponzielli, Shili Duan, Steve McMahon, Michael D.Cole, Linda Z. Penn, and Cheryl H. Arrowsmith , Characterizing cMyc MBII and its interaction with TRRAP, submitted

Appendix 5: Romina Ponzielli, Paul C. Boutros, Sigal Katz, Angelina Stojanova, Adam P. Hanley, Fereshteh Khosravi, Igor Jurisica, Linda Z. Penn. Optimization of experimental design parameters for high-throughput chromatin immunoprecipitation studies, submitted

Supporting Data

Figures 1A, 1B, 2A, 2B

Research Article

The c-Myc Oncogene Directly Induces the H19 Noncoding RNA by Allele-Specific Binding to Potentiate Tumorigenesis

Dalia Barsyte-Lovejoy,[1] Suzanne K. Lau,[3,4] Paul C. Boutros,[1,2,4] Fereshteh Khosravi,[1] Igor Jurisica,[2,4,6] Irene L. Andrulis,[7] Ming S. Tsao,[3,4,5] and Linda Z. Penn[1,4]

Divisions of [1]Cancer Genomics and Proteomics, [2]Signaling Biology, and [3]Applied Molecular Oncology, Ontario Cancer Institute/Princess Margaret Hospital; Departments of [4]Medical Biophysics and [5]Laboratory Medicine and Pathobiology, [6]Computer Science, University of Toronto, Canada; and [7]Fred Litwin Cancer Genetics Center, Samuel Lunenfeld Research Institute Mount Sinai Hospital, Toronto, Ontario, Canada

Abstract

The product of the *MYC* oncogene is widely deregulated in cancer and functions as a regulator of gene transcription. Despite an extensive profile of regulated genes, the transcriptional targets of c-Myc essential for transformation remain unclear. In this study, we show that c-Myc significantly induces the expression of the *H19* noncoding RNA in diverse cell types, including breast epithelial, glioblastoma, and fibroblast cells. c-Myc binds to evolutionarily conserved E-boxes near the imprinting control region to facilitate histone acetylation and transcriptional initiation of the *H19* promoter. In addition, c-Myc down-regulates the expression of insulin-like growth factor 2 (*IGF2*), the reciprocally imprinted gene at the *H19/IGF2* locus. We show that c-Myc regulates these two genes independently and does not affect *H19* imprinting. Indeed, allele-specific chromatin immunoprecipitation and expression analyses indicate that c-Myc binds and drives the expression of only the maternal *H19* allele. The role of H19 in transformation is addressed using a knockdown approach and shows that down-regulation of H19 significantly decreases breast and lung cancer cell clonogenicity and anchorage-independent growth. In addition, c-Myc and H19 expression shows strong association in primary breast and lung carcinomas. This work indicates that c-Myc induction of the *H19* gene product holds an important role in transformation. (Cancer Res 2006; 66(10): 5330-7)

Introduction

The transforming members of the Myc family (c-Myc, N-Myc, and L-Myc) show deregulated expression in a broad spectrum of cancers, including carcinomas of the lung, breast, and prostate as well as leukemias and lymphomas (1). c-Myc is a transcription factor that, with its obligate heterodimerization partner Max, binds to DNA sequence elements called E-boxes (2). c-Myc-Max can subsequently recruit histone acetyltransferase (HAT) activity (3), chromatin remodeling complexes (4), or promote RNA polymerase II (RNAPII) clearance (5) to allow for target gene transcription. c-Myc-Max can also repress gene transcription primarily by interfering with the assembly or function of the transcriptional complex (6–8). As a transcription factor, c-Myc regulates numerous gene targets that subsequently execute its many biological activities, including cell proliferation, transformation, angiogenesis, and apoptosis (9). Identifying these target genes is key in elucidating the role of this potent oncogene in transformation and has thus received much attention. Despite an extensive list of c-Myc-regulated genes, it remains unclear which cohort of target genes is responsible for the strong transforming activity of c-Myc (10, 11).

Recent analyses using advanced high-throughput chromatin immunoprecipitation technology has revealed the nature of the target genes whose promoter regulatory regions are bound by c-Myc (12–15). Remarkably, these analyses have indicated that c-Myc target genes include both coding and noncoding RNAs (ncRNA; ref. 16). Noncoding RNAs are transcripts expressed and processed in the nucleus in a manner similar to protein coding genes; however, ncRNAs lack a conserved open reading frame. Although elucidating the function of ncRNAs is in the early stages of investigation, evidence suggests that at least some may have roles in tumorigenesis (17). To determine the role of c-Myc-regulated ncRNAs in transformation, we investigated both the regulation and function of the large prototypic ncRNA, H19, as a downstream target of c-Myc.

H19 was first described as a tumor suppressor (18, 19), but more recent analysis shows that H19 expression is reactivated in breast (20), endometrial (21), lung (22), cervical (23), esophageal (24), and bladder (25) tumors. The *H19*/insulin-like growth factor 2 (*IGF2*) locus, containing both the *H19* and *IGF2* genes, is subject to genomic imprinting, which leads to differential allelic expression of H19 from the maternal allele and IGF2 from the paternal allele (26). This allele-specific expression is highly regulated by differential methylation of CpG dinucleotides that are usually concentrated in CpG islands, genomic elements that are often located close to promoter regions (27). As our data indicated that c-Myc can bind to intergenic regions containing CpG islands (13), we also explored the consequences of allele-specific CpG methylation on the transcriptional regulatory function of c-Myc at the *H19/IGF2* locus.

In this study, we show that c-Myc induces the expression of the H19 ncRNA and binds directly to E-boxes close to the imprinting control region (ICR). Using allele-specific chromatin immunoprecipitation analysis, we show that c-Myc specifically binds and regulates the active maternal *H19* allele and does not bind or affect the expression of the silenced paternal allele. In addition, c-Myc down-regulates transcription of the reciprocally imprinted gene *IGF2*. The significance of H19 up-regulation by c-Myc and the association of c-Myc and H19 transcript levels were assessed in primary and established tumor cells derived from breast and lung cancer patients.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Linda Z. Penn, Division of Cancer Genomics and Proteomics, Ontario Cancer Institute/Princess Margaret Hospital, 610 University Avenue, Toronto, Canada M5G 2M9. Phone: 416-946-2276; Fax: 416-946-2840; E-mail: lpenn@uhnres.utoronto.ca.

doi:10.1158/0008-5472.CAN-06-0037

Materials and Methods

Cell lines. The immortal, nontransformed MCF10A breast cell line (gift from Dr. Muthuswamy) was cultured as described (28). Cells were grown with 10% fetal bovine serum in α-MEM (glioblastoma T98G), DMEM H21 (rat cardiomyocyte H9C2 and fibroblast Rat1MycER^{TAM}), McCoy's (breast cancer cell lines MDA-MB231, SKBR3, and colon carcinoma line HCT116), and RPMI 1640 (T47D breast cancer cell line and lung cancer cell lines A549, H460, and H520). Rat Myc null cells (HO15.19) were grown in DMEM H21 with 10% calf serum. Where indicated, 5 μmol/L 5-aza-deoxycytidine (AzaC; Sigma, St. Louis, MO) was added to cells every 24 hours. Trichostatin A (TSA; Calbiochem, La Jolla, CA) was used at 300 nmol/L.

Retroviral gene transfer. Ectopic human c-Myc was introduced by infection with ecotropic, replication-incompetent retrovirus, and expression was confirmed as described (29).

Isolation and analysis of RNA. Total RNA was isolated as previously described (30) and purified using the RNeasy Mini kit (Qiagen, Inc., Chatsworth, CA). Five micrograms of total cellular RNA were reverse transcribed using Superscript II Reverse Transcription reagents and OligoD_T (Invitrogen, San Diego, CA). Northern blots were done as described (11).

Gene expression analysis in cancer cell lines. Semiquantitative reverse transcription-PCR (RT-PCR) was carried out as described (13) with primers provided in Supplementary Table S1, and the conditions are available upon request. For the quantitative real-time RT-PCR analysis of the breast and lung cell lines, SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA) were used according to the manufacturer's instructions. A total of one thousandth of the cDNA reaction was used for each PCR triplicate. The results were normalized to the levels of the 36B4 transcript using the comparative C_t method. The allele-specific PCR employed the forward primers specific for the particular *H19* allele, CGGCTCTCGAAGGTGAAGCT (B) or CGGCTCTCGAAGGTGAAGCG (A), and the reverse primer used was TCGTGGAGGCTTTGAATCTCTCAG.

Selection of patients. The RNA expression profiles of a total of 186 breast cancer samples, representing 137 distinct tumors from the cohort described (31), were assayed using cDNA expression microarrays. Two samples lacked follow-up data and were excluded from further analysis. A total of 240 snap-frozen non–small cell lung (NSCLC) carcinoma samples were harvested from patients who have been treated primarily by surgical resection at the University Health Network from 1996 to 2000. Tissues were banked with informed consent, and the studies have been approved by the institutional Research Ethics Board.

Analysis of breast tumor expression data. Raw microarray images were quantitated with the GenePix (Axon, Union City, CA) software package with flagging of low-quality spots. Spot signal was calculated by subtracting the median background pixel intensity from the mean foreground pixel intensity. Ratios were transformed into $\log_2$ space, and missing values were imputed from spotwise duplicates where possible. Data were normalized by sequential print-tip loess smoothing (within array) and scale adjustment (between arrays) as described by Yang et al. (32). All normalization employed the limma package of the Bioconductor library for R (v 2.0.1; ref. 33). Normalized expression values were centered using a 15% trimmed mean. To assess the spot quality, we determined the correlation of the two sequences representing H19 (BI092679 and BQ028553) and MYC (H43827 and W87741). Five outliers with extreme *H19* expression were identified via the *Q* test and removed, leading to higher spot correlation ($R = 0.72$). Spots were then collapsed by averaging. Similarly, the expression from two MYC spots were found to be largely uncorrelated ($R = -0.07$); thus, the clone most specific to MYC was selected. These normalized data represented 137 distinct tumor samples, with clinical follow-up available for 135 of these. These samples were dichotomized around the median H19 expression level. A *t* test with the assumption of unequal variances was used to test MYC expression for significant differences. Normalized array data for all spots are available as Supplementary Table S2.

Quantification and analysis of *MYC* and *H19* expression in lung tumor samples. Real-time quantitative PCR amplification was conducted using the SYBR Green assay in the ABI PRISM 7900-HT (Applied Biosystems). Each 10-μL quantitative RT-PCR reaction contained a 2-ng equivalent of cDNA in a 384-well plate. The reactions were activated at 95°C for 3 minutes followed by 40 cycles of 95°C for 15 seconds, 65°C for 15 seconds, and 72°C for 20 seconds. The transcript number/ng cDNA was obtained using standard curves generated with a pool of 10 nontumor lung genomic DNAs. Primer sequences are provided in Supplementary Table S1. Duplicate RT samples were used in each assay. Technical replicates displayed high correlation ($R_{avg} = 0.96 \pm 0.04$; Supplementary Fig. S1) and were then collapsed through averaging. First, expression values were $\log_2$ transformed after addition of a pseudocount. Samples lacking H19 or TBP signal were removed from the data set. For each of the remaining samples, a normalization factor was calculated using the mean of the four housekeeping genes (*TBP, ACTB, B2M*, and *BAT1*) and used to remove nonbiologicalvariability.Forthe*i*thpatient,thenormalizationfactoris N_i $N_i = 0.25 \times [X_i(\text{TBP}) + X_i(\text{ACTB}) + X_i(\text{B2M}) + X_i(\text{BAT1})]$, and normalized expression values for gene M are $Y_i(M) = X_i(M) - N_i$, where the X_i values correspond to unnormalized expression values. For each patient cohort, the normalized expression values were median centered to yield the final expression estimates: $Z_i(M) = Y_i(M) - \text{median}[Y(M)]$. The two patient cohorts were then merged, and the overall data set of 240 distinct samples was dichotomized around the median H19 expression level. A *t* test with the assumption of unequal variances was used to test for differential c-Myc expression. Raw quantitative RT-PCR data are available online as Supplementary Table S3.

Chromatin immunoprecipitation. Chromatin immunoprecipitation was done as previously described (13) using the following antibodies: 2 μg Myc (Sc-764), 1 μg CTCF, 0.5 μg RNAPII (Sc-764), 1 μg Max, all from Santa Cruz Biotechnology (Santa Cruz, CA), and 0.5 μg of AcH3 or AcH4 (Upstate Biotechnology, Lake Placid, NY). Real-time PCR was done as described above using human genomic DNA as the standard and normalizing the specific antibody signal to the input signal. Primer sequences are provided in the Supplementary Table S1, and the conditions are available upon request. The same chromatin immunoprecipitation material was used for allele specific PCR with allele-specific primers CGCCTACTTATGTGATGATCAG or CGCCTACTTATCTGATGATCAC and the reverse GCACCCACGATAATGGATT.

H19 knockdown. The small interfering RNAs (siRNA) were designed using the SFold web site (http://sfold.wadsworth.org). The control siRNA was against luciferase (34), whereas H19 hairpin oligo sequences are CCGGGCGGGTCTGTTTCTTTACTTTCAAGAGAAGTAAAGAAACAGACC-CGCTTTTTG and reverse AATTCAAAAAGCGGGTCTGTTTCTTTACTTCT-CTTGAAGTAAAGAAACAGACCCGC. The annealed phosphorylated oligos were cloned into pLKO1puro lentiviral vector (gift from Drs. Stewart, Novina, and Weinberg). These constructs, together with packaging vectors pMD.G, pMDLg/pRRE and pRsv-Rev (gift from Dr. Naldini), were transfected into 293TV cells, and viral supernatant was collected 48 hours later and used to infect the cells.

Anchorage and clonogenicity assays. Anchorage-independent growth assays were done as described (29), except 5,000 cells were seeded and counted at the end of a 2-week period. For clonogenicity assays, 500 cells were seeded in six-well dishes, and the media containing puromycin was changed every 3 days. After the 1- to 2-week period, the resulting colonies were stained with 2% methylene blue in 50% ethanol and counted.

Results

c-Myc up-regulates H19 and down-regulates IGF2 transcripts. The *H19/IGF2* locus is subject to genomic imprinting (Fig. 1*A*). Allele-specific methylation of CpG dinucleotides in the ICR leads to H19 expression from the maternal allele, whereas the reciprocally imprinted *IGF2* gene is expressed from the paternal allele (26). The ICR on the maternal allele is unmethylated and bound by CTCF, a zinc-finger protein that acts as a boundary between the enhancers located 3′ of H19 and the promoters of IGF2. The paternal allele is methylated at the ICR, preventing

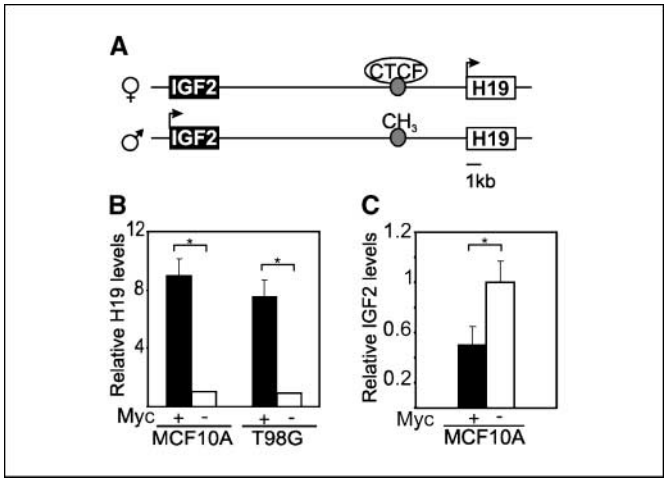


Figure 1. Ectopic Myc induces the expression of H19 and represses IGF2 expression. *A,* schematic of the *H19/IGF2* locus, where active promoters are denoted with arrows and the imprinting control region (*gray circles*) is shown bound by CTCF or methylated (CH_3). *B,* H19 ncRNA expression in MCF10A and T98G cells with (+) and without (−) ectopic Myc as assessed by quantitative RT-PCR. 36B4 ribosomal protein mRNA was used as a normalization control. *C,* IGF2 mRNA levels in MCF10A cells. Quantitative RT-PCR was conducted in triplicates twice. *Columns,* mean; *bars,* SD. *, $P < 0.05$, statistical significance as assessed using a paired *t* test.

CTCF binding, allowing the enhancers to potentiate IGF2 transcription (26).

To evaluate whether c-Myc regulates *H19* expression, we introduced ectopic c-Myc into several cell types, including MCF10A immortalized nontransformed mammary epithelial cells and T98G glioblastoma cells (Supplementary Fig. S2A). Cells with ectopic c-Myc expression showed 7- to 10-fold up-regulation of H19 ncRNA expression as assessed by real-time quantitative RT-PCR (Fig. 1*B*), semiquantitative RT-PCR, and Northern blotting (Supplementary Fig. S2B). This up-regulation was also evident in diploid fibroblasts WI38 (data not shown) and a medulloblastoma cell line UW228 (see Fig. 3*B*). The activation of the constitutively expressed c-Myc/estrogen receptor regulatory region ($MycER^{TAM}$) chimera by 4-hydroxytamoxifen also resulted in elevated H19 expression in Rat-1 fibroblasts (Supplementary Fig. S2C).

Given the strong up-regulation of H19, we assessed the expression of IGF2 in response to exogenous c-Myc expression. Interestingly, c-Myc down-regulated IGF2 in several cell systems, such as MCF10A and WI38 (Fig. 1*C*; data not shown). However, in T98G cells, which have low to undetectable levels of IGF2, down-regulation of IGF2 was not consistently detectable (data not shown). This prompted further evaluation in rat cardiomyocytes that have high basal levels of IGF2 transcripts. c-Myc robustly repressed the levels of IGF2 in H9C2 cardiomyocytes as shown by Northern blot (Supplementary Fig. S2D). Thus, c-Myc strongly up-regulates H19 and down-regulates IGF2 transcript levels in several cell types.

c-Myc directly binds to the regulatory regions of *H19* and *IGF2*. To determine whether c-Myc directly regulates H19, we assessed *in vivo* genomic DNA binding of c-Myc to the regulatory region of H19 using chromatin immunoprecipitation, focusing on the evolutionary conserved E-boxes situated 1.5 and 3.1 kb upstream from the transcription start site. c-Myc immunoprecipitates were highly enriched in these DNA fragments compared with control serum immunoprecipitates (Fig. 2*A*, primer sets 1 and 3). The ICR region containing the second cluster of three DNase hypersensitive sites essential for CTCF binding showed weaker binding of c-Myc (Fig. 2*A*, primer set 2). Interestingly, the abovementioned E-boxes are 650 and 360 bp away from the second hypersensitive site cluster within the ICR, and there are also numerous noncanonical E-boxes interspersed between the

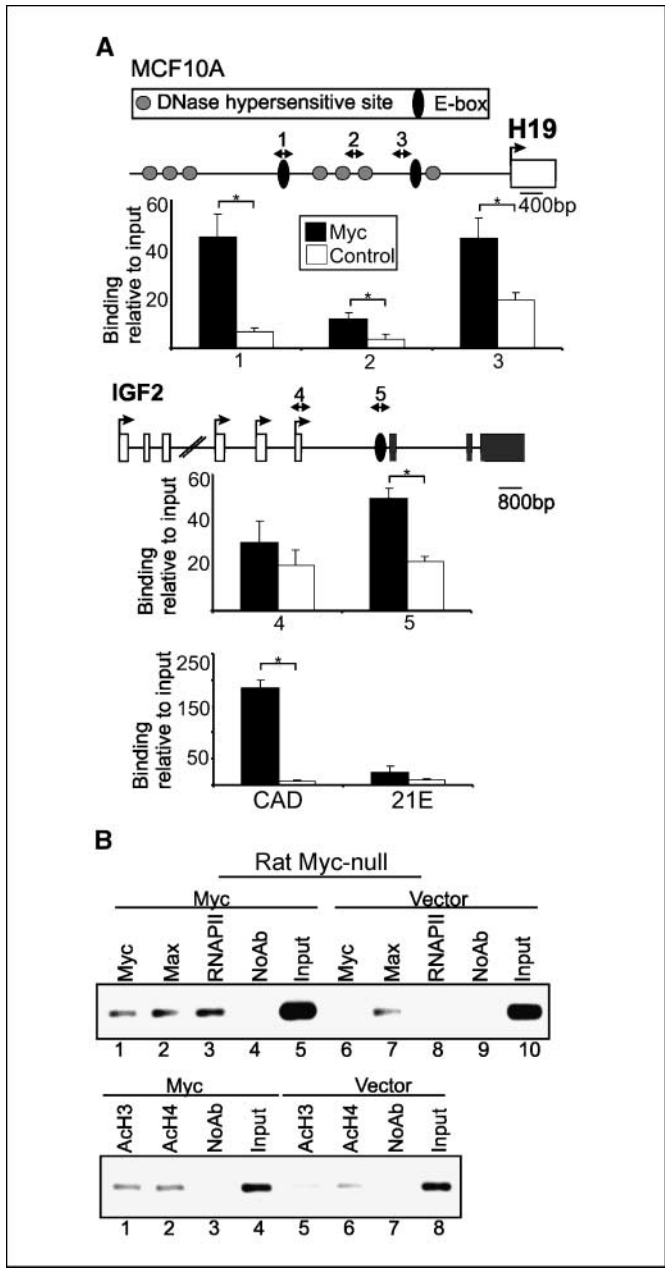


Figure 2. Myc binds to the promoter regions of H19 and IGF2 to regulate gene transcription. *A,* Myc binding at the regulatory region of the *H19* and *IGF2* genes, as well as the CAD (positive control) and chromosome 21 E-box (negative control) was assessed by chromatin immunoprecipitation using Myc-specific antibody (*Myc*) or preimmune serum (*Control*) in MCF10A cells expressing ectopic Myc. Primers are indicated by arrows. DNase-hypersensitive sites (*gray circles*) comprise the ICR. Relative binding was normalized to the input signal and expressed as 0.001% of input. Quantitative PCR was conducted in triplicates twice. *Columns,* mean; *bars,* SD. *, $P < 0.05$, statistical significance as assessed using a paired *t* test. *B,* the transcriptional components bound at the *H19* promoter were assessed by chromatin immunoprecipitation using primers to the proximal E-box (equivalent to primer set 3 in *A*). Rat Myc-null cells that were reconstituted with Myc (*Myc*) or control green fluorescent protein vector (*vector*) were assayed using indicated antibodies. Chromatin immunoprecipitation experiments were repeated at least twice with representative gels shown.

hypersensitive sites. Therefore, from the binding data alone it remained unclear whether Myc would regulate the promoter of H19 or the ICR and imprinting. This issue is addressed through functional analysis below. We have also investigated c-Myc binding to the *IGF2* gene using a scanning chromatin immunoprecipitation approach. *IGF2* is expressed from the promoter P4 in MCF10A cells (data not shown); however, c-Myc binding was weak or undetectable not only at the P4 promoter but also at several other sites, including the differentially methylated region 1 and the coding region in the exon 9 (Fig. 2*A*, primer set 4; data not shown). Only the E-box 5′ from coding exon 7 showed binding of c-Myc (Fig. 2*A*, primer set 5). The positive control *CAD* promoter was highly enriched in c-Myc immunoprecipitates, whereas the negative control E-box at chromosome 21 showed no significant enrichment (Fig. 2*A*). Thus, c-Myc binds to the regulatory regions of *H19* and *IGF2 in vivo*.

c-Myc recruits HATs to the *H19* promoter. To investigate evolutionary conservation and further assess the mechanism of *H19* transcriptional regulation by c-Myc, we compared rat fibroblast cells that are devoid of Myc expression with those that have been reconstituted with ectopic c-Myc for promoter occupancy by chromatin immunoprecipitation analysis. In the absence of c-Myc, Max is bound to the *H19* promoter proximal E-box, a feature previously described as a hallmark of c-Myc-regulated genes (Fig. 2*B, top, lane 7*; ref. 13). Reintroducing c-Myc into these cells resulted in binding of c-Myc, the recruitment of RNAPII (Fig. 2*B, top, lanes 1* and *3*) and an increase in the acetylation of histone H3 (AcH3) and H4 (AcH4) at the *H19* E-box (Fig. 2*B, bottom, lanes 1* and *2*). The role of histone acetylation in c-Myc-mediated transcriptional induction of *H19* was further supported by evaluating the effect of the histone deacetylase (HDAC) inhibitor TSA. Treatment with TSA resulted in increased expression of H19 in $myc^{-/-}$ cells as detected by semiquantitative RT-PCR (Fig. 3*A, lanes 1-4*). Curiously, the level of *H19* expression achieved by TSA treatment was higher in $myc^{-/-}$ cells than in c-Myc-expressing cells (Fig. 3*A, lanes 3* and *4*). Thus, histone acetylation activity is recruited by c-Myc to the *H19* promoter, leading to its activation.

c-Myc does not affect imprinting of the *H19/IGF2* locus. Because the paternal allele of *H19* is usually silent due to ICR methylation (see Fig. 1*A*), the effect of DNA methylation on c-Myc regulation of H19 and IGF2 was further investigated. *H19* expression was evaluated in $myc^{-/-}$ cells and c-Myc reconstituted cells in the presence and absence of DNA methylation inhibitor AzaC. Exposure to AzaC did not result in H19 up-regulation in $myc^{-/-}$ cells (Fig. 3*A*, compare *lanes 1, 5*, and *7*), whereas blocking DNA methylation increased *H19* expression in c-Myc-reconstituted cells (Fig. 3*A*, compare *lane 2* with *lanes 6* and *8*). By removing DNA methylation (AzaC) and then increasing histone acetylation (TSA), *H19* expression was potentiated in $myc^{-/-}$ cells (Fig. 3*A*, compare *lanes 1, 3*, and *9*), and no further induction by c-Myc was evident under these conditions (Fig. 3*A, lanes 9* and *10*). Blocking DNA methylation also did not affect IGF2 basal gene expression (Fig. 3*A*, compare *lanes 1, 5*, and *7*) but further potentiated c-Myc repression of this target gene (Fig. 3*A*, compare *lanes 2, 6*, and *8*). c-Myc repression of *IGF2* remains intact despite treatment with both TSA and AzaC (Fig. 3*A, lanes 9* and *10*). These data suggest that c-Myc regulation of the *H19/IGF2* locus does not involve DNA methylation or imprinting. Moreover, loss of methylation alone is insufficient to alter basal gene expression but can cooperate with c-Myc to potentiate the regulation of these target genes.

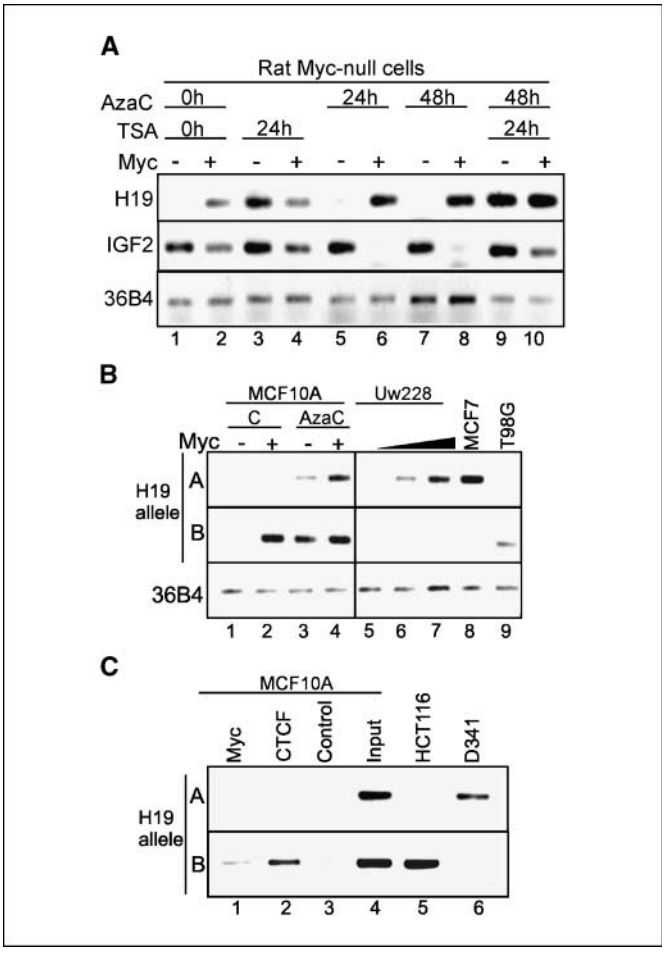


Figure 3. Myc binds and regulates only one allele of H19. *A,* semiquantitative RT-PCR of H19 and IGF2 in Myc null rat fibroblasts expressing the control vector (−) or ectopic Myc (+) that were exposed to HDAC inhibitor TSA and/or DNA methylation inhibitor AzaC for the time indicated. The ribosomal protein 36B4 mRNA serves as a control. Semiquantitative RT-PCR was conducted twice with similar results. *B,* allelic expression analysis of H19 using semiquantitative RT-PCR. *C,* allele-specific binding by Myc and CTCF in MCF10A cells that were used for chromatin immunoprecipitation with Myc-specific and CTCF-specific antibodies, and the immunoprecipitates were analyzed with allele-specific primers. PCR analysis was conducted thrice; representative data are shown.

c-Myc up-regulates only one allele of *H19*. Because the *H19/IGF2* locus is subject to genomic imprinting, we had a unique opportunity to investigate the transcriptional regulatory effect of c-Myc on the allelic expression of *H19*. Multiple cell lines were first tested for polymorphisms in the *H19* locus (data not shown), and the allele-specific primers A and B (*Alu*I polymorphism) were then designed, differing by a 3′ single nucleotide. The primers were tested on the homozygous MCF7 and T98G cells to show allelic specificity (Fig. 3*B, lanes 8* and *9*). In MCF10A cells, ectopic expression of c-Myc induced expression of *H19* from the single allele B (Fig. 3*B, lanes 1* and *2*). c-Myc induction of both alleles of H19 was evident in the presence of AzaC, showing that in the absence of DNA methylation, c-Myc can access and activate both alleles of *H19* (Fig. 3*B, lanes 3* and *4*). Similarly, ectopic expression of increasing amounts of c-Myc activates only the single allele A of *H19* in a dose-dependent manner in Uw228 medulloblastoma cells (Fig. 3*B, lanes 5-7*). Therefore, c-Myc induces *H19* expression by a mechanism that is restricted to only one allele of *H19*.

c-Myc binds to the promoter of one *H19* allele. To further evaluate the molecular mechanism of c-Myc induced allele-specific expression of *H19,* we queried whether c-Myc binds to only one or both alleles of *H19 in vivo*. Allele-specific chromatin immunoprecipitation primers were tested on control DNA from homozygous HCT116 and D341 cell lines to ensure that only one allele is recognized using this approach (Fig. 3*C, lanes 5* and *6*). Both c-Myc and CTCF are bound to allele B but not allele A in MCF10A cells (Fig. 3*C*, compare *lanes 1-3* and *lanes 2-3*). The weaker signal for c-Myc binding could be due to the allele-specific primers being situated 0.4 kb away from the E-box. The bound allele was designated as allele B (Fig. 3*B*) because it has been reported that CTCF binds to the maternally derived, expressed allele of *H19* (26). Indeed, allele B of *H19* is bound by c-Myc and CTCF (Fig. 3*C, lanes 1* and *2*), whereas allele A is not bound by either c-Myc or CTCF, nor it is sensitive to induced expression in response to ectopic c-Myc in MCF10A cells. These allelic binding results were also confirmed by methyl-specific PCR (data not shown). Thus, c-Myc and CTCF bind to the maternally derived, expressed allele of *H19*.

H19 knockdown inhibits tumorigenic properties of breast and lung cancer cells. To assess the role of H19 in transformation, we evaluated the expression and functional significance of H19 in cells derived from breast and lung carcinomas. H19 has recently been reported to be elevated in a high proportion of these tumor types (20, 22). H19 expression was evident in the SKBR3 and T47D breast cancer cell lines and the lung adenocarcinoma A549 cells, whereas expression was undetectable in MDA-MB231 breast cancer cell line (Fig. 4*A*). Stable siRNA-mediated knockdown of H19 resulted in significant decrease in both clonogenicity, as assessed by the efficacy of colony formation on solid support (Fig. 4*B*), and anchorage-independent growth, as assessed by colony formation in soft agar (Fig. 4*C*). The inhibition of both clonogenicity- and anchorage-independent growth in response to H19 knockdown was significant in all cell lines with H19 expression (SKBR3, T47D, and A549), and insignificant in cells with undetectable basal H19 expression, such as MDA-MB231 cells or H460 and H520 lung carcinoma cell lines (Fig. 4*B* and *C*; data not shown). A lack of biological effect of H19 siRNA in MDA-MB231, H460, and H520 serves as an important specificity control. Three of five siRNAs designed worked well to down-regulate H19 expression; however, only one was specific (data not shown). Another siRNA we identified as a nonspecific H19 siRNA was previously used to transiently down-regulate H19 (35). Thus, H19 contributes to the clonogenic and anchorage-independent growth properties in breast and lung cancer cells with reactivated H19 expression.

H19 regulation of IGF2. As H19 has been reported to repress the levels of IGF2, we also monitored the effect of H19 knockdown on IGF2 levels. No significant and consistent increase in IGF2 expression upon H19 repression was noted in any of the breast cancer cells (Fig. 4*D*) or MCF10A cells expressing ectopic c-Myc and elevated H19 (data not shown). Only the A549 cells with H19 knockdown displayed increased IGF2 transcript expression (Fig. 4*D*).

c-Myc binds to the *H19* promoter in SKBR3 but not MDA-MB231 cells. To further evaluate the mechanism of c-Myc induction of *H19* transcription, we queried whether c-Myc was bound to the *H19* E-box (−1.5 kb) in cells with (SKBR3) and without (MDA-MB231) basal *H19* expression. Quantitative chromatin immunoprecipitation analysis showed that c-Myc, CTCF, RNAPII, and AcH4 were present at the *H19* E-box in SKBR3 compared with control antibody (Fig. 5*A*). However, in MDA-MB231 cells, binding of c-Myc and RNAPII was not evident, whereas the binding of CTCF was reduced, and AcH4 was unaffected (Fig. 5*A*). The positive control E-box (CAD) bound c-Myc and RNAPII in both cell lines, whereas the negative control E-box on chromosome 21 had no significant binding of any of the factors (Fig. 5*B* and *C*). Therefore, c-Myc binds to the promoter region of *H19* in SKBR3 cells and contributes to the up-regulation of H19.

High expression levels of H19 correlate with elevated levels of c-Myc mRNA in breast and lung cancer patients. To determine if c-Myc regulation of H19 is a feature of primary human

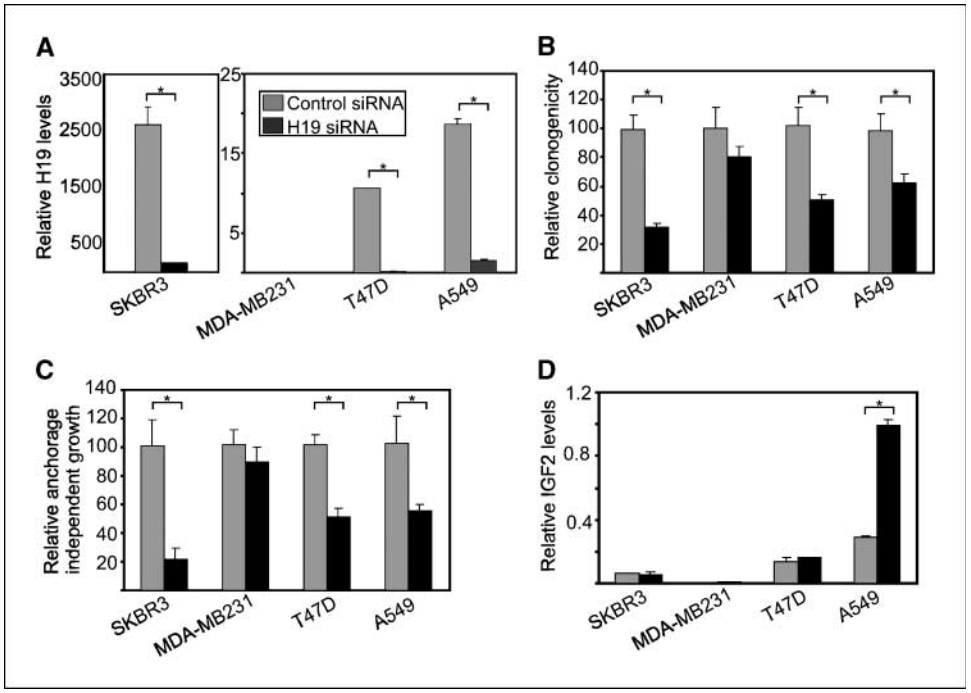


Figure 4. H19 siRNA inhibits cellular transformation. *A,* quantitative RT-PCR assessment of the basal H19 expression and efficacy of H19 siRNA. 36B4 mRNA was used for normalization. The graph of SKBR3 is illustrated separately due to the very high H19 expression. Quantitative RT-PCR was conducted twice in triplicate. *Columns,* mean; *bars,* SD. *, $P < 0.05$, statistical significance as assessed using a paired *t* test. *B,* clonogenicity is decreased in response to the ectopic expression of H19 siRNA in cell lines with basal H19 expression. The experiment was conducted two independent times with consistent results; a representative experiment is shown. Experimental results were normalized between the cell lines. *Columns,* mean; *bars,* SD. *, $P < 0.05$, statistical significance as assessed using a paired *t* test. *C,* anchorage-independent growth is decreased in response to H19 siRNA. Data were analyzed, and significance test was done as in (*B*). *D,* quantitative RT-PCR analysis of IGF2 expression was conducted as in *A*.

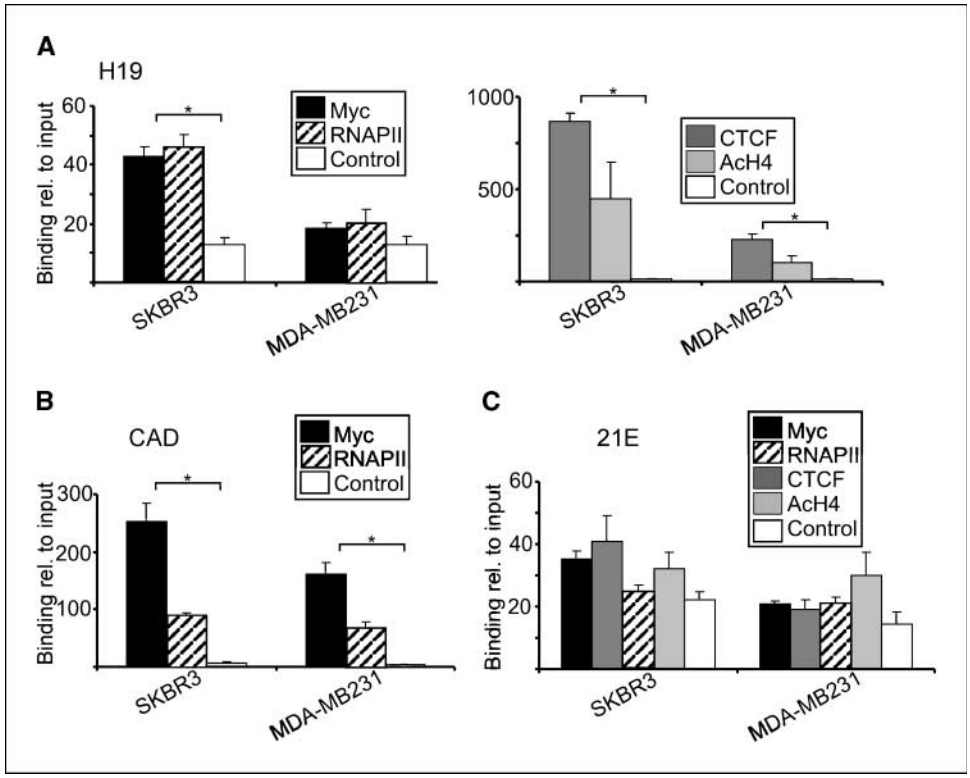


Figure 5. Transcriptional complex binding at the *H19* promoter in SKBR3 and MDA-MB231 cells. *A,* chromatin immunoprecipitation and quantitative PCR analysis of transcriptional components binding the *H19* E-box. *Left,* Myc and RNAPII binding relative to control; *right,* CTCF and AcH4 binding. *B,* chromatin immunoprecipitation and quantitative PCR analysis of Myc and RNAPII binding at the positive control CAD in both cell lines. *C,* chromatin immunoprecipitation and quantitative PCR analysis of negative control E-box binding on chromosome 21. Real-time quantitative PCRs were done and analyzed as in Fig. 2. *, $P < 0.001$, statistical significance as assessed using a *t* test.

tumors, we obtained expression estimates for both genes from a large microarray study of 137 node-negative breast cancer cases. Following data normalization and centering, patients were dichotomized into two equal-sized groups based on H19 expression levels (Fig. 6). Using a two-tailed heteroscedastic *t* test, we show that c-Myc expression was significantly higher ($P = 0.009$) in breast tumors with high H19 levels than those showing lower H19 expression. To show that this is a feature of other tumor types, we characterized the relationship between H19 and c-Myc expression in lung cancers using quantitative RT-PCR in a panel of 240 NSCLCs. Following normalization to a battery of four housekeeping genes, we again dichotomized patients into two equal-sized groups based on H19 expression levels (Fig. 6). Using a two-tailed heteroscedastic *t* test, we again found that c-Myc expression was significantly higher ($P = 0.002$) in tumors with elevated H19 levels than those showing lower H19 expression. To ensure that our results are not artifacts of the statistical analysis, we verified the results using a two-log threshold with the hypergeometric test and again found statistical significance ($P = 0.01$). Both boxplot and spikeplot representations of the data are available as Supplementary Figs. S3 and S4. These analyses show positive correlation of high H19 levels with elevated c-Myc mRNA levels. Selected clinical information is available in Supplementary Tables S4 and S5.

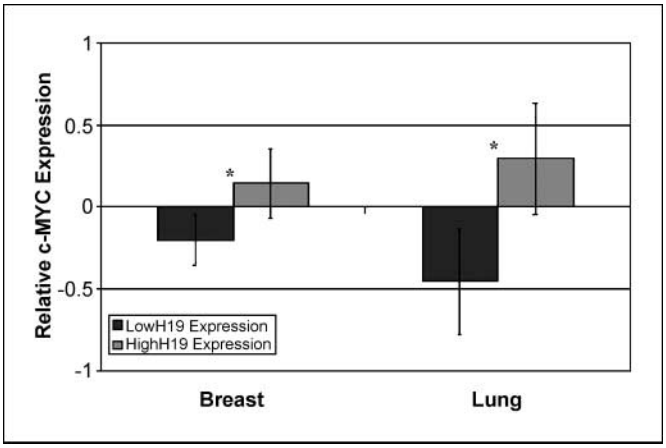


Figure 6. The expression of H19 and c-Myc are related in both breast ($n = 137$) and lung ($n = 240$) tumor samples. For each tumor type, tumors were dichotomized into two equal-sized groups based on H19 expression, and c-Myc expression was compared between the high and low H19 cohorts. *, $P < 0.01$, significance by a two-tailed *t* test. *Columns,* mean; *bars,* 95% confidence intervals.

Discussion

Based on allelic expression and genomic binding studies, we show that c-Myc directly binds to the *H19* promoter and highly up-regulates the transcription of the maternal *H19* allele by recruiting HAT activity. We further show that H19 knockdown, in a panel of breast and lung cancer cell lines, results in the reduction of their tumorigenic phenotype as shown by foci formation and anchorage-independent growth assays. Indeed, a strong association between c-Myc and H19 transcript levels was evident in both primary breast and lung cancer patient material. Taken together, these results indicate that Myc up-regulation of H19 strongly contributes to the tumorigenic phenotype of breast and lung cancer cells.

Our study shows that c-Myc binds to the conserved E-boxes at the *H19* promoter close to the ICR and up-regulates the expression of this ncRNA in MCF10A breast epithelial and established rat fibroblast cell lines. The physical boundary between the ICR and the promoter of *H19* remains undefined; thus, we investigated the mechanism of regulation and show that Myc does not affect ICR

function or imprinting of this locus, as H19 remains expressed from one allele only. In a parallel study, Lee et al. also found that c-Myc up-regulates H19 in mouse liver *in vivo*.[8] However, their data indicate the strongest binding of Myc at an enhancer downstream of the *H19* promoter in hepatocytes. We did not see significant binding of Myc to this region in MCF10A cells expressing ectopic Myc (data not shown). The distinct binding sites and possibly induction mechanisms used by Myc in human and rat transformed cells compared with murine nontransformed hepatocytes is intriguing and likely is a consequence of cell type differences. Importantly, only the maternal allele of *H19* was regulated by Myc in both studies.

Our results show that c-Myc binds and regulates the maternal *H19* allele strongly supporting the observation that the c-Myc/Max complex is unable to bind methylated E-boxes *in vitro* (36). While this article was in preparation, N-Myc was also reported to selectively bind to unmethylated E-boxes (37). This suggests that E-box methylation is a determinant of c-Myc and N-Myc as regulators of gene transcription. This is further supported by our results showing that the silenced paternal allele did not display binding or regulation by Myc in MCF10A cells, whereas removal of DNA methylation leads to induction of the paternal allele by c-Myc. These data suggest that changes in the DNA methylation patterns, which are common in cancer (38), could alter transcriptional profiles and thus modulate the biological function of Myc.

Our results indicating that c-Myc binds and activates *H19* in an allele-specific manner suggests that additional transcription factors, as well as Myc, may regulate transcription with allelic preference. Recent analyses have shown that allelic variation in gene expression is common even in the nonimprinted genes in mammalian genomes and may affect 20% to 50% of human genes (39). We also show that the main function of Myc at the *H19* promoter is to recruit HAT activity and RNAPII for transcriptional initiation. In cells lacking Myc, basal H19 expression was induced only in response to HDAC inhibition but not to the inhibition of DNA methylation. Similarly, the silencing of $p16^{INK4A}$ can also occur in the absence of DNA methylation through histone modifications (40). Therefore, Myc binds the ICR of *H19* in an allele-specific manner to induce transcription through the recruitment of HAT activity.

Furthermore c-Myc down-regulates IGF2 transcripts in cell lines with detectable IGF2 mRNA. Mechanistically, our data indicate that Myc binding is restricted to the maternal allele ICR and is unlikely to have an effect on IGF2 expression from the paternal allele. Although it has been reported that H19 can repress IGF2 expression (41), we did not consistently observe this cross-regulation in the cell systems studied, suggesting this potential regulatory mechanism may be a cell-dependent and/or tumor type–dependent phenomenon. Surprisingly, our data indicate that Myc binds the E-box in the first intron of *IGF2* to repress transcription. Most Myc repressed genes analyzed to date are regulated through initiator or proximal promoter regions (6–8). It remains to be determined whether E-box-dependent Myc repression is evident and unique to IGF2 or functional at multiple loci. Moreover, it remains unclear whether this potentially novel mechanism is best captured through genomic analysis *in vivo* and not detectable by conventional transient indicator gene analysis methods. Biologically, gene expression analysis of a series of neuroblastoma cell lines indicated that high N-Myc levels correlate with low IGF2 levels (42). However, colon carcinomas display high levels of Myc and loss of *IGF2* imprinting (expressed from both alleles) that is sometimes accompanied by elevated levels of IGF2 (43). This implies that the repression of IGF2 is cell specific or is lost in colon cancer cells. At this time, the significance of IGF2 down-regulation by Myc remains unclear. As a mitogen in many cell types, IGF2 is part of a signaling cascade that is able to induce Myc and Myc repression of IGF2 could function as part of negative feedback loop to control mitogen stimulation. A similar function has been proposed for Myc repression of the platelet-derived growth factor-β receptor (7).

The physiologic role of H19 ncRNA seems to be restricted to the time of embryonic expression (44), and its pathologic role has only recently been investigated. The introduction of H19 into some cell lines resulted in anchorage-independent growth suppression, thus earning H19 tumor suppressor designation (18, 19). Recently, however, H19 expression has been shown to be reactivated in a variety of tumors (20–25). Ectopic expression of H19 in MDA-MB231 and T24P cells increased their growth and tumorigenicity (45, 46), whereas the subsequent knockdown of H19 in MDA-MB231 reduced cell proliferation (35). We show that the knockdown of H19 in SKBR3, T47D, and A549 cells leads to the reduction of their clonogenic ability and decreased anchorage-independent growth. The analysis of H19 and c-Myc expression in primary breast and lung tumor samples indicated that high H19 levels are strongly associated with high Myc transcript levels. This strong association in primary tissue, in combination with the essential role of H19 in transformation, suggests that Myc-induced H19 expression contributes to both tumor etiology and Myc's function as an oncogene.

The function of H19 remains enigmatic. Although knockout *H19* mice seem grossly normal (47), we and others show that H19 plays a role in the tumorigenic phenotype (35, 45, 46). Based on the physical association of the H19 transcript with polysomes, the mechanism of H19 action is thought to be at the level of translational regulation (48). In addition, thioredoxin, a modulator of signal transduction and potentiator of tumorigenesis, was recently identified as translationally up-regulated by H19 (49). Further studies will undoubtedly elucidate other H19-regulated molecules and the role of H19 in both physiologic and pathologic settings. As more transcriptional regulators of H19 are determined, H19 regulation and function will also be better understood. Although this work was in progress, E2F1 was shown to regulate *H19* (35), which suggests that Myc and E2F1 can exert a positive combinatorial control over H19 transcription. It is notable that genes bound by Myc (12–15) and E2F1 (50) display a considerable overlap.

We have thus identified *H19* as a Myc-up-regulated gene that potentiates the tumorigenic phenotype of breast and lung cancer cells. Complex interactions between DNA methylation and histone modifications regulate H19 expression, where the role of Myc is to recruit HAT activity to unmethylated E-boxes and initiate allele-specific *H19* transcription.

Acknowledgments

Received 1/6/2006; revised 2/15/2006; accepted 3/9/2006.

Grant support: National Cancer Institute of Canada with funds from the Canadian Cancer Society (L.Z. Penn), Terry Fox Run (I.L. Andrulis), Canadian Institutes of Health Research Scholarships (P.C. Boutros and D. Barsyte-Lovejoy), Institute for Robotics and Intelligent Systems (I. Jurisica). In addition, this research was also supported by the U.S. Department of Defense Breast Cancer Research Program grant BC032138 (L.Z. Penn), as such, the content of the information does

[8] Personal communication.

not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

A special thank you to Drs. Linda Lee and Chi Dang (Johns Hopkins University, Baltimore, MD) for communicating unpublished data in advance of publication. We thank Dr. Muthuswamy (Harvard, Boston, MA) for MCF10A cells; Dr. Huang (Hospital for Sick Kids, Canada) for Uw228 cell and their Myc containing derivative cDNA; Drs. Stewart, Novina, and Weinberg (Whitehead, MA) for the lentiviral transfer plasmid; Dr. Naldini (San Raffaele Telethon Institute, Italy) for lentiviral packaging constructs. We especailly indebted to Drs. Mao, Oster, and Katz for technical assistance and S. Colby for breast cancer expression array analysis. We thank the Penn laboratory for constructive discussion and review of the manuscript.

References

1. Nesbit CE, Tersak JM, Prochownik EV. MYC oncogenes and human neoplastic disease. Oncogene 1999;18:3004–16.

2. Oster SK, Ho CS, Soucie EL, et al. The myc oncogene: MarvelouslY Complex. Adv Cancer Res 2002;84:81–154.

3. McMahon SB, Van Buskirk HA, Dugan KA, et al. The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. Cell 1998;94:363–74.

4. Cheng SW, Davies KP, Yung E, et al. c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function. Nat Genet 1999;22:102–5.

5. Eberhardy SR, Farnham PJ. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. J Biol Chem 2002;277:40156–62.

6. Barsyte-Lovejoy D, Mao DY, Penn LZ. c-Myc represses the proximal promoters of GADD45a and GADD153 by a post-RNA polymerase II recruitment mechanism. Oncogene 2004;23:3481–6.

7. Mao DY, Barsyte-Lovejoy D, Ho CS, et al. Promoter-binding and repression of PDGFRB by c-Myc are separable activities. Nucleic Acids Res 2004;32:3462–8.

8. Staller P, Peukert K, Kiermaier A, et al. Repression of $p15^{INK4b}$ expression by Myc through association with Miz-1. Nat Cell Biol 2001;3:392–9.

9. Dang CV. c-Myc target genes involved in cell growth, apoptosis, and metabolism. Mol Cell Biol 1999;19:1–11.

10. O'Connell BC, Cheung AF, Simkevich CP, et al. A large scale genetic analysis of c-Myc-regulated gene expression patterns. J Biol Chem 2003;278:12563–73.

11. Watson JD, Oster SK, Shago M, et al. Identifying genes regulated in a Myc-dependent manner. J Biol Chem 2002;277:36921–30.

12. Orian A, van Steensel B, Delrow J, et al. Genomic binding by the Drosophila Myc, Max, Mad/Mnt transcription factor network. Genes Dev 2003;17:1101–14.

13. Mao DY, Watson JD, Yan PS, et al. Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression. Curr Biol 2003;13:882–6.

14. Li Z, Van Calcar S, Qu C, et al. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proc Natl Acad Sci U S A 2003;100:8164–9.

15. Fernandez PC, Frank SR, Wang L, et al. Genomic targets of the human c-Myc protein. Genes Dev 2003;17:1115–29.

16. Cawley S, Bekiranov S, Ng HH, et al. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 2004;116:499–509.

17. Costa FF. Non-coding RNAs: new players in eukaryotic biology. Gene 2005;357:83–94.

18. Hao Y, Crenshaw T, Moulton T, et al. Tumour-suppressor activity of H19 RNA. Nature 1993;365:764–7.

19. Isfort RJ, Cody DB, Kerckaert GA, et al. Role of the H19 gene in Syrian hamster embryo cell tumorigenicity. Mol Carcinog 1997;20:189–93.

20. Adriaenssens E, Dumont L, Lottin S, et al. H19 overexpression in breast adenocarcinoma stromal cells is associated with tumor values and steroid receptor status but independent of p53 and Ki-67 expression. Am J Pathol 1998;153:1597–607.

21. Tanos V, Ariel I, Prus D, et al. H19 and IGF2 gene expression in human normal, hyperplastic, and malignant endometrium. Int J Gynecol Cancer 2004;14:521–5.

22. Kondo M, Suzuki H, Ueda R, et al. Frequent loss of imprinting of the H19 gene is often associated with its overexpression in human lung cancers. Oncogene 1995;10:1193–8.

23. Douc-Rasy S, Barrois M, Fogel S, et al. High incidence of loss of heterozygosity and abnormal imprinting of H19 and IGF2 genes in invasive cervical carcinomas. Uncoupling of H19 and IGF2 expression and biallelic hypomethylation of H19. Oncogene 1996;12:423–30.

24. Hibi K, Nakamura H, Hirai A, et al. Loss of H19 imprinting in esophageal cancer. Cancer Res 1996;56:480–2.

25. Ariel I, Sughayer M, Fellig Y, et al. The imprinted H19 gene is a marker of early recurrence in human bladder carcinoma. Mol Pathol 2000;53:320–3.

26. Reik W, Walter J. Genomic imprinting: parental influence on the genome. Nat Rev Genet 2001;2:21–32.

27. Macleod D, Ali RR, Bird A. An alternative promoter in the mouse major histocompatibility complex class II I-Abeta gene: implications for the origin of CpG islands. Mol Cell Biol 1998;18:4433–43.

28. Debnath J, Muthuswamy SK, Brugge JS. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. Methods 2003;30:256–68.

29. Oster SK, Mao DY, Kennedy J, et al. Functional analysis of the N-terminal domain of the Myc oncoprotein. Oncogene 2003;22:1998–2010.

30. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 1987;162:156–9.

31. Andrulis IL, Bull SB, Blackstein ME, et al. neu/erbB-2 amplification identifies a poor-prognosis group of women with node-negative breast cancer. Toronto Breast Cancer Study Group. J Clin Oncol 1998;16:1340–9.

32. Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 2002;30:e15.

33. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 2003;3:1–26.

34. Paddison PJ, Caudy AA, Bernstein E, et al. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes Dev 2002;16:948–58.

35. Berteaux N, Lottin S, Monte D, et al. H19 mRNA-like noncoding RNA promotes breast cancer cell proliferation through positive control by E2F1. J Biol Chem 2005;280:29625–36.

36. Prendergast GC, Ziff EB. Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. Science 1991;251:186–9.

37. Perini G, Diolaiti D, Porro A, et al. *In vivo* transcriptional regulation of N-Myc target genes is controlled by E-box methylation. Proc Natl Acad Sci U S A 2005;102:12117–22.

38. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nat Rev Genet 2002;3:415–28.

39. Yan H, Yuan W, Velculescu VE, et al. Allelic variation in human gene expression. Science 2002;297:1143.

40. Bachman KE, Park BH, Rhee I, et al. Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. Cancer Cell 2003;3:89–95.

41. Wilkin F, Paquette J, Ledru E, et al. H19 sense and antisense transgenes modify insulin-like growth factor-II mRNA levels. Eur J Biochem 2000;267:4020–7.

42. Raetz EA, Kim MK, Moos P, et al. Identification of genes that are regulated transcriptionally by Myc in childhood tumors. Cancer 2003;98:841–53.

43. Kaneda A, Feinberg AP. Loss of imprinting of IGF2: a common epigenetic modifier of intestinal tumor risk. Cancer Res 2005;65:11236–40.

44. Brannan CI, Dees EC, Ingram RS, et al. The product of the H19 gene may function as an RNA. Mol Cell Biol 1990;10:28–36.

45. Lottin S, Adriaenssens E, Dupressoir T, et al. Overexpression of an ectopic H19 gene enhances the tumorigenic properties of breast cancer cells. Carcinogenesis 2002;23:1885–95.

46. Ayesh S, Matouk I, Schneider T, et al. Possible physiological role of H19 RNA. Mol Carcinog 2002;35:63–74.

47. Jones BK, Levorse JM, Tilghman SM. Igf2 imprinting does not require its own DNA methylation or H19 RNA. Genes Dev 1998;12:2200–7.

48. Li YM, Franklin G, Cui HM, et al. The H19 transcript is associated with polysomes and may regulate IGF2 expression in trans. J Biol Chem 1998;273:28247–52.

49. Lottin S, Vercoutter-Edouart AS, Adriaenssens E, et al. Thioredoxin post-transcriptional regulation by H19 provides a new function to mRNA-like non-coding RNA. Oncogene 2002;21:1625–31.

50. Weinmann AS, Yan PS, Oberley MJ, et al. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev 2002;16:235–44.

Trial watch

MANY USES FOR LAPATINIB

At the recent American Society of Clinical Oncology annual conference, positive clinical trial results with lapatinib in various cancer types were presented. Lapatinib is a dual kinase inhibitor that targets the epidermal-growth-factor receptors 1 (EGFR or ERBB1) and 2 (HER2 or ERBB2).

A phase III open-label study reported by Charles Geyer was designed to evaluate time to progression (TTP) in women with ERBB2-positive trastuzumab-resistant advanced breast cancer treated with lapatinib plus capecitabine, or capecitabine alone. Of 321 women evaluated, the median TTP in the combination group was 36.9 weeks, compared with 19.7 weeks in the monotherapy group. Median progression-free survival was 36.9 weeks on combination therapy versus 17.9 weeks on capecitabine alone. Survival data are not yet mature, but are similar so far. The superior efficacy of the combination therapy has led the Independent Data Monitoring Committee to recommend discontinuing enrolment into the trial.

In addition, Neil Spector reported the results of an ongoing phase II study of lapatinib monotherapy for women with relapsed or refractory inflammatory breast cancer. 62% of ERBB2 overexpressors had a complete or partial response; there were no responders in the group who overexpressed ERBB1 but not ERBB2. These data are based on only 57 patients, but they indicate that ERBB1 expression alone cannot predict sensitivity to lapatinib.

Another phase III trial was conducted in patients with advanced renal-cell carcinoma (RCC) — a tumour known to produce high levels of ERBB1. In this study, 417 patients with immunotherapy-refractory RCC were randomized to receive standard second-line hormonal therapy or lapatinib. Alain Ravaud reported that while overall survival and TTP were similar, those 241 patients who greatly overexpressed ERBB1 had longer TTP if on lapatinib rather than hormonal therapy (15.1 weeks versus 10.9 weeks), as well as overall survival (46 weeks versus 37.9 weeks). Follow-up in this trial is continuing.

WEB SITE http://www.asco.org

FOCUSING ON AFRICAN-AMERICAN WOMEN

African-American women with breast cancer are known to have a poorer prognosis than non-African-American women. A case–control study reported in *JAMA* is unique in that it oversampled African-American and pre-menopausal women to enable better evaluation of these populations in terms of tumour subtype and survival than has previously been possible.

496 incident cases of invasive breast cancer were included and subtyped into luminal, basal-like, ERBB2 positive and oestrogen-receptor negative (ER^-), and unclassified. Basal-like breast cancer was the most prevalent subtype among pre-menopausal African-American women (39%) compared with postmenopausal African-American women (14%) or non-African-American women of any age (16%). Conversely, the luminal subtype was less prevalent. There was no difference within the $ERBB2^+ER^-$ subtype with regards to race or menopausal status. The overall disease-specific survival was 80%, but the shortest survival was among the $ERBB2^+ER^-$ (52%) and basal-like subtypes (75%). These findings could contribute to the poor prognosis of young African-American women with breast cancer, but additional studies of long-term survival in other African-American populations are required to confirm this.

ORIGINAL RESEARCH PAPER Carey, L. A. *et al.* Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* **295**, 2492–2502 (2006)

TUMORIGENESIS

The silent messenger

Various genes are known to be regulated by the transcription factor MYC, but the genes that are crucial for its tumorigenic qualities are not clearly defined. Linda Penn and colleagues have found that a non-coding RNA, H19, is a transcriptional target of MYC, and the inhibition of *H19* transcription in cancer cells suppresses certain cellular characteristics that are associated with oncogenic transformation.

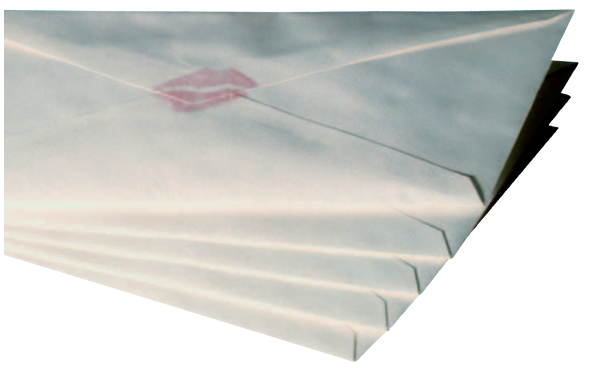
Despite being first identified as a tumour suppressor, increased expression of H19 is seen in various human cancers, including lung and breast cancer. *H19* resides at the same locus as the insulin-like growth factor 2 gene (*IGF2*), and both of these genes are subject to genomic imprinting — only the maternal allele of *H19* and the paternal allele of *IGF2* are expressed. Through the combined use of reverse transcriptase (RT)-PCR, northern blotting and allele-specific chromatin immunoprecipitation techniques the authors were able to show that MYC binds to the regulatory regions of the *H19* and *IGF2* genes. Furthermore, MYC only binds to the non-methylated promoter of the maternal allele of *H19* to induce transcription, which indicates that MYC does not alter the imprinting of this gene and that MYC target genes are determined by promoter methylation.

Small interfering RNAs that target *H19* in cells derived from lung and breast tumours inhibited anchorage-independent growth and colony formation *in vitro* — two well-known characteristics of transformed cells. But is there a correlation between the level of MYC expression and H19 expression in these tumours? The authors undertook a large microarray study using 137 node-negative breast cancer samples. Statistical analyses showed that expression levels of *MYC* were significantly higher in samples with high levels of *H19* expression. Similar results were obtained using 240 non-small-cell lung cancer samples, which were analysed using real-time RT-PCR.

The function of non-coding RNAs is unclear, but it is thought that the *H19* transcript might regulate RNA translation. So, MYC might indirectly regulate the expression of a number of proteins in tumour cells through the increased expression of H19.

Nicola McCarthy

ORIGINAL RESEARCH PAPER Barsyte-Lovejoy, D. *et al.* The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer Res.* **66**, 5330–5337 (2006)



ELSEVIER

European Journal of Cancer 41 (2005) 2485–2501

European Journal of Cancer

www.ejconline.com

Cancer therapeutics: Targeting the dark side of Myc

Romina Ponzielli [1], Sigal Katz [1], Dalia Barsyte-Lovejoy [1], Linda Z. Penn *

Division of Cancer Genomics and Proteomics, Ontario Cancer Institute/Princess Margaret Hospital, Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, Ont., Canada M5G 2M9

Available online 20 October 2005

Abstract

The potent Myc oncoprotein plays a pivotal role as a regulator of tumorigenesis in numerous human cancers of diverse origin. Experimental evidence shows that inhibiting Myc significantly halts tumour cell growth and proliferation. This review summarises recent progress in understanding the function of Myc as a transcription factor, with emphasis on key protein interactions and target gene regulation. In addition, major advances in drug development aimed at eliminating Myc are described, including antisense and triple helix forming oligonucleotides, porphyrins and siRNA. Future anti-Myc strategies are also discussed that inhibit Myc at the level of expression and/or function. Targeting the dark side of Myc with novel therapeutic agents promises to have a profound impact in combating cancer.

Keywords: Myc; Max; Therapeutics; Deregulation; Transformation; Target genes; Protein-protein interactors; siRNA; Antisense oligonucleotides; ChIP-chip; Cancer

1. Introduction

These are exciting times for clinicians and scientists working at the front line of cancer research. In recent years, our understanding of the genetic etiology and molecular dependence of cancer has advanced to the point where we can begin to exploit this knowledge for the design of novel, effective, anti-cancer therapeutics. Several new classes of anti-cancer agents have been developed to target pathways that are essential for cancer cell growth and survival. To date, the effort has been primarily focused at targeting growth factor receptors which has led to achievements, such as Gleevec (Imatinib, STI-571), a small molecular weight kinase inhibitor, and Trastuzumab (Herceptin), an antibody reagent that targets the HER-2/neu receptor [1–4]. These agents highlight the merits of molecular targeting and provide an important proof-of-concept that targeted therapeutics can be successfully developed and applied to patient care. However, activating mutations can occur at multiple independent points along oncogenic signalling cascades, so targeting cell surface receptors restricts the range of potentially sensitive tumours. Downstream transcription factors that directly control the transformation program provide an alternative target that may be activated in a broader spectrum of cancers. The Myc transcription factor is one of the most potent and frequently deregulated oncoproteins in human cancers [5,6]. Multiple extracellular and intracellular signalling cascades converge to regulate the Myc oncogene (Fig. 1) making it an especially attractive target in the control of transformation.

In this review, we focus on the most recent advances in understanding the molecular mechanisms of Myc function in the etiology of human cancers. Special emphasis is placed on Myc interactions with other proteins and its gene-regulatory mechanisms. As well, we provide an overview of the past, present and potential future strategies used to target Myc.

* Corresponding author. Tel.: +1 416 946 2276; fax: +1 416 946 2840.
E-mail address: lpenn@uhnres.utoronto.ca (L.Z. Penn).
[1] These authors contributed equally to the work.

0959-8049/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2005.08.017

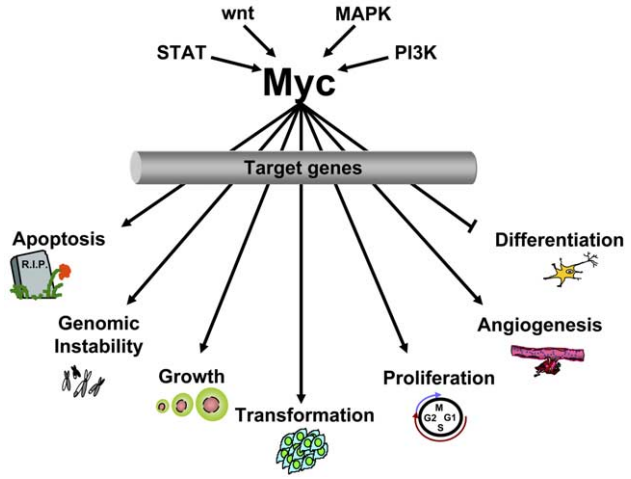


Fig. 1. The regulatory network of Myc. Myc deregulation is a hallmark of many cancers and occurs as a consequence of activation of one or more signalling pathways that induce Myc expression and function as a regulator of gene transcription. These include mitogen activated protein kinases (MAPK) [192–194], phosphatidylinositol-3 kinase (PI3K) [195,196], wnt-TCF/LEF pathway [197], and signal transducer and activator of transcription (STAT) [198] pathways. The target genes regulated by Myc orchestrate the many biological activities attributed to Myc, including apoptosis, genomic instability, growth, transformation, proliferation, angiogenesis and blocking differentiation.

1.1. Myc expression and biological activities

Members of the *myc* family that show oncogenic activity in human cancers include *c-myc*, *MYCN* and *MYCL1*. In a normal physiological state, the protein products of all three genes are expressed during fetal development, whereas only c-Myc protein is expressed in adult tissues. c-Myc protein levels, in non-transformed cells, are induced or suppressed by virtually all signalling cascades bearing proliferative and anti-proliferative cues, respectively. Mitogen stimulation induces c-Myc as an immediate-early response gene, whose expression is essential and sufficient for G1/S progression [7–9]. c-Myc also plays a role in G2/M transition, making it one of the key players in cell cycle regulation [10]. As such, it is important that *c-myc* mRNA and protein have a very short half-life (20–30 min) and are tightly regulated. c-Myc expression is normally rapidly responsive to environmental cues and has been dubbed, "the intracellular sentinel of the extracellular milieu" [11]. By contrast to the highly regulated state of c-Myc and the absence of N-myc and L-myc expression in normal cells, cancer cells often harbour deregulated expression of any one of these three *myc* oncogenes [11].

The founding member of the family, *c-myc*, was first shown to be an oncogene when it was identified as the transduced *v-myc* gene of the transforming avian myelocytomatosis retrovirus [12]. Deregulated c-Myc expression was subsequently shown to be prolific in human cancers. In 100% of Burkitt's lymphoma, c-Myc is translocated with an immunoglobulin enhancer that drives high levels of constitutive *c-myc* mRNA and protein expression, which is instrumental in initiating the disease [13]. Translocations involving the c-Myc locus have also been reported in several additional tumours, including diffuse large cell lymphoma, T-cell acute lymphocytic leukaemia and multiple myeloma [14]. Amplification of c-Myc and/or deregulated expression is evident in many tumours including melanomas and carcinomas of the breast, prostate and colon [5,6,14]. Amplification of *MYCN*, is a hallmark of neuroblastoma [15,16], while *MYCL1* is amplified in ovarian cancer [17]. All three transforming members of the *myc* family can be amplified in non-small cell lung carcinoma [14]. Importantly, in recent years it has become clear that deregulation is not restricted to gross genetic abnormalities of the *myc* gene family, such as translocation or amplification, but can also occur as a consequence of direct or indirect mutations of regulatory molecules controlling *myc* gene expression [11]. Thus, deregulation of Myc expression is evident in numerous human cancers of diverse origin and can result from mutations at one or multiple levels of regulation. Unless otherwise stated, Myc will refer to data described for c-Myc, but these results are often relevant for the highly similar N-Myc and L-Myc oncoproteins.

In response to signals from the cellular environment, Myc can regulate a broad variety of distinct biological activities (Fig. 1). In addition to Myc driving cell proliferation, growth, and transformation, deregulated Myc has also been shown to increase apoptosis, genomic instability, and angiogenesis as well as block differentiation [18–28]. The prevailing model is that Myc controls such a disparate set of activities by regulating distinct cohorts of target genes that then orchestrate each activity. One of the major gaps in the field is linking Myc function, as a transcriptional regulator, to the wide-range of biological activities controlled by Myc [29]. In the following sections we provide an overview of the enormous effort from several labs to identify the target genes regulated by Myc and to uncover the protein–protein interactions that are essential for Myc to function as a transcription factor in transformation.

1.2. Myc structure

Traditionally the Myc protein has been divided into an N-terminal domain (NTD) involved in transactivation and transrepression; and a C-terminal domain (CTD) that is critical for DNA binding. Myc is a transcription factor of the basic helix-loop-helix leucine zipper (b-HLH-LZ) superfamily [30]. The C-terminus harbours both the primary nuclear localisation signal and the basic motif required for binding to the CACGTG E-box DNA recognition sequence (Fig. 2)

[31,32]. The HLH-LZ domain, essential for all known Myc activities, is the heterodimerisation domain that is required for Myc to bind to its primary partner protein, Max (Myc associated protein X).

The N-terminus is a major regulatory region responsible for assembly of the transcriptional machinery [11]. Within the N-terminus there are several highly conserved sequences termed Myc boxes which, together with the C-terminal b-HLH-LZ, define the Myc family of proteins (Fig. 2). Although Myc box I (MBI) is required for gene activation, deletion of this region only partially abolishes the transforming ability of Myc [11,33]. Myc box II (MBII) is essential for the ability of Myc to transform, drive cell proliferation, inhibit differentiation, repress gene transcription, and activate certain target genes [11]. Recently, a third conserved region of Myc has been described, Myc Box III (MBIII), that plays a role in transformation, lymphomagenesis and apoptosis [34]. Interestingly, recent results using circular dichroism indicated that the N-terminal domain shows little to no inherent secondary structure, suggesting that protein–protein interactions are essential for proper folding and function of Myc [35,36]. From a therapeutic perspective this observation offers hope that specific inhibitors can disrupt unique points of interaction between Myc and its binding partners thereby inhibiting transformation. This may offer a novel approach to targeting oncogenic Myc in a tumour-specific manner.

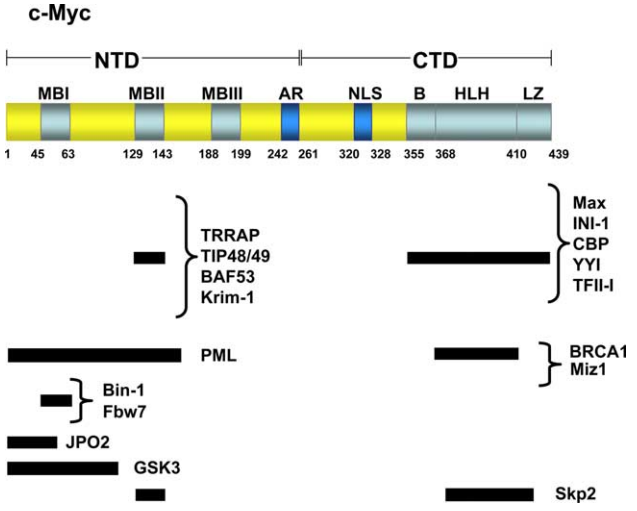


Fig. 2. The structural domains of human c-Myc and their link to protein–protein interactions. c-Myc contains at least six regions which are highly conserved between Myc paralogs and orthologs. The Myc N-terminal domain (NTD) is defined as amino acids 1–262 and contains Myc Box I (MBI), Myc Box II (MBII) and Myc Box III (MBIII) and the acidic region (AR). The Myc C-terminal domain (CTD) is defined as residues 263–439 and contains the primary nuclear localisation signal (NLS) and the basic helix-loop-helix leucine zipper domains (B-HLH-LZ). The regions of c-Myc necessary for the interaction of the specific proteins are shown.

2. Myc interacting proteins

For Myc to hold its extensive role in the control of cellular function, a network of key binding proteins is required (Fig. 2). The identification of this network started with the discovery of Max over a decade ago by screening a human cDNA library with a radiolabeled fusion protein containing the Myc CTD [37]. In recent years, several additional Myc-binding proteins have primarily been identified using biochemical and two-hybrid methodologies. The Myc–Max interaction is already being explored as a therapeutic target, while inhibitors to other key interacting proteins more recently identified, such as TRRAP, have yet to be investigated. A new array of interactors has been identified in recent years, but their role in Myc dependent transformation and the effect of their disruption needs to be further evaluated before they are justifiable targets for therapeutic intervention.

2.1. Myc and Max: from structure to biology

Myc activity is entirely dependent upon dimerisation with Max, an abundant, ubiquitously expressed b-HLH-LZ protein. Recently, the co-crystal structure has been solved and shows that homodimers of Max are held together by polar interactions alone, while Myc–Max heterodimers are stabilised by charged interactions [38]. This results in Myc–Max dimers forming more easily than Max-Max dimers. The crystal structure shows that Myc–Max dimers are stabilised by hydrophobic and polar/charged interactions via helices in the LZ region. Two positively charged residues in Myc form a tetrad with Max, and these two pairs of hydrogen bonds alone control heterodimer specificity with Max. This difference between charged and polar residues also explains the disfavour of Myc–Myc homodimers, caused by electrostatic repulsions between the complementary residues [38,39]. Perhaps one of the most interesting findings from the crystallographic data is the formation of a tetramer with two Myc–Max heterodimers oriented head to tail of the LZ, each binding a DNA E-box. It has been proposed that biologically they may bind widely separated E-box sequences; however, the crystal structure reveals no specific hydrogen bond interactions to stabilise this tetramer [38]. Another function of the heterotetramer might be the formation of a platform for assembly of additional protein factors such as Miz-1, INI1 and BRCA1 which bind to the b-HLH-LZ region of Myc (Fig. 2). The detailed structure of Myc–Max may give rise to novel therapeutic strategies that interfere with either the formation of the tetramer or the association with Myc CTD interactors.

2.2. Myc interaction with chromatin remodeling proteins

In addition to the well-documented association with Max, Myc has been shown to interact with a number of additional transcription factors and co-factors that modulate its activity.

The DNA bound Myc–Max heterodimer interacts through the Myc N-terminal region with a variety of proteins involved in transcription. TRRAP is of particular interest as it forms part of a multiprotein complex with histone acetyl-transferase (HAT) activity. Myc contributes to chromatin remodeling through an MBII-dependent interaction with TRRAP [40,41]. Inhibition of TRRAP synthesis or function blocks Myc-mediated oncogenesis, establishing an essential role for TRRAP in Myc activity [42]. The yeast homologue of TRRAP, tra1 is a component of the SAGA (SPT/ADA/GCN5/Acetyltransferase) complex [43] which also contains the HAT GCN5. The human homologue, hGCN5, has been shown to interact in a complex with Myc through TRRAP [44]. A Myc-Gcn5 fusion protein can partially restore the ability of a Myc MBII deletion mutant to induce transformation, suggesting that recruitment of HAT activity is an important component of Myc function. Myc also interacts with the p400 complex containing TRRAP, surprisingly this complex lacks HAT activity suggesting that Myc-TRRAP interaction may serve additional roles in addition to recruiting HATs. Nevertheless, p400 function in Myc-mediated oncogenesis remains unclear [45]. The recent discovery that CBP (CREB binding protein) binds to the Myc CTD provides an additional link between Myc CTD and activation of transcription. CBP functions partly through its HAT activity and was shown to interact with Myc *in vivo* and to stimulate Myc dependent transactivation. Interestingly, CBP also acetylates Myc *in vitro* and co-expression of the two proteins resulted in stabilisation of Myc [46]. Functionally, the biological significance of this interaction, in particular for Myc-mediated transformation, is yet to be determined.

Myc interacts with several other proteins implicated in chromatin remodeling, namely TIP48 and TIP49 [47]. Interaction of these molecules with Myc requires the Myc NTD and occurs independently of TRRAP binding. TIP48 and TIP49 have ATP hydrolysing activity, as well as suspected helicase activity and have been shown to be required for the foci formation by Myc and Ras in a primary co-transformation assay [47].

Myc may also be involved in a second mechanism of chromatin remodeling: an ATP-dependent process involving the recruitment of the SWI/SNF complex, which regulates transcription through nucleosome repositioning. Myc directly interacts with INI1/hSNF5, a key component of the SWI/SNF complex, and this was suggested to stimulate Myc transcriptional activity [48]. INI1/hSNF5 appears to be a tumour suppressor, and is mutated in the majority of atypical teratoid and malignant rhabdoid tumours [49–51]. More recently, Myc has been shown to interact with BAF53, an actin-related protein which is another integral component of the SWI/SNF chromatin remodeling complex. Targeted mutations in BAF53 inhibit oncogenic transformation by Myc [52]. In that context, it will be important to determine if other proteins of the SWI/SNF complex have the potential to interact with Myc and whether SWI/SNF recruitment may play a role in Myc-mediated transformation. Despite this enormous progress, many questions remain. Which genes are targeted by each of these distinct complexes? Does each complex drive a specific genetic program, such as transformation or apoptosis? Clearly, the precise role of these ATPase/helicase family proteins in Myc biology requires further study and this knowledge will be instructive in determining which complex is of highest priority for therapeutic targeting to block Myc transforming function.

2.3. Interactions with other transcriptional regulators

Whereas TRRAP appears to be a positive regulator of Myc-mediated transformation, Bin-1 (Bridging protein-1) appears to be a negative regulator [53]. The mechanism by which Bin-1 represses transformation is not fully understood, but it includes effects on the cell cycle as well as the promotion of apoptosis in response to Myc over-expression [54]. Since Bin-1 is deleted in a variety of tumours we sought to better define its interaction with Myc. Our group has used nuclear magnetic resonance (NMR) spectroscopy and biochemical assays to define the mechanism of interaction between Bin-1 and the MYC NTD [55]. We showed that a small proline-rich peptide within the conserved MBI interacts with the SH3 domain of Bin1 and that this interaction can be disrupted by phosphorylation of Myc Ser62. Our findings raise the intriguing possibility that the conserved MBI region may bind to other SH3 domain proteins. In addition, the data highlights the significance of post-translational modifications to Myc activity and suggests their modulation as an opportunity for therapeutic intervention.

Due to the evidence that Myc NTD binding proteins are critical to Myc function, we developed a novel high throughput screen termed the repressed transactivator assay (RTA) [56]. This two hybrid approach enables the full Myc NTD to be used as bait in the context of Myc as a transactivator bound to DNA. Novel methods, such as the RTA, will greatly facilitate the identification of Myc-binding proteins and inhibitors to disrupt these interactions. Using the RTA we screened a library derived from medulloblastoma cells with the Myc NTD as bait. A novel Myc interactor, JPO2, was isolated and shown to be a transcription factor containing a putative

LZ and a ring finger domain. JPO2 is closely related to a Myc transcriptional target, JPO1, and was also recently identified as a novel member of this emerging JPO family [57–59]. We show that JPO2 expression affects anchorage independent growth and is associated with metastasis in medulloblastomas [57]. As we have also observed Myc interaction with JPO1 protein, an interesting working model at a molecular level that emerges, sees Myc inducing JPO1 transcription for the formation of a Myc–JPO complex as a positive feedback loop to regulate gene transcription. It will be interesting to determine the validity of this model for the JPO family and to determine if additional Myc induced genes can also form functional transcription regulatory complexes with Myc.

Another novel Myc NTD binding protein is Krim-1, a nuclear zinc finger protein that contains a KRAB domain. Krim-1 was identified using the Ras Recruitment System (RRS) and was shown to associate with MBII [60,61]. In a reporter assay, Krim-1 was able to negatively regulate Myc transactivation and inhibit its oncogenic activity in REFs, a phenotype that is reduced in the presence of a Myc-binding mutant of Krim-1 [61].

Another interactor is the ARF tumour suppressor protein, which binds directly to Myc and inhibits its transcriptional activity in a p53-independent fashion. ARF blocks Myc ability to activate transcription without affecting its ability to repress transcription. ARF prevents Myc-induced transformation whereas Myc-induced apoptosis remains intact even in the absence of p53. ARF was shown to bind both Myc NTD and HLH-LZ domains of Myc [62,63]. These findings suggest a safeguard mechanism for preventing aberrant Myc signalling. The question naturally follows: does ARF differentially bind to a cohort of Myc-target genes related to proliferation and/or transformation? Studies using specific inhibitors of the Myc-ARF interaction will provide a direct link between ARF suppression of tumourigenesis and the control of Myc transcriptional activities. These findings highlight the critical role of MBII in transcriptional regulation and suggest that complexes, such as TRRAP, may either compete or cooperate with other multiprotein complexes for binding with ARF or Krim1.

Recently, several additional Myc-binding proteins have been described, but their role in transformation remains unclear. For example, Myc and PML were shown to co-localise within discrete nuclear structures associated with the nuclear matrix, termed PML bodies [64,65]. These PML bodies associate with regions of high transcriptional activity in the genome and have been implicated in diverse cellular processes, although their specific function remains open [66]. The differential effect that PML appears to have on Myc activation is also interesting. In PML-null mouse embryonic fibroblasts (MEFs) the expression levels of numerous Myc target genes is altered [64,65]. Although PML was demonstrated to physically interact with the Myc NTD and is recruited at Myc regulated promoters, the exact mechanism by which PML influences Myc activity remains unclear. Further insight into the precise contribution of these interactors, as well as others such as PARP-10 [67] or human papillomavirus E6 [68], to the function of oncogenic Myc in the carcinogenic process will determine whether targeting these proteins may be of merit.

2.4. Interactors that regulate Myc stability

Multiple mechanisms ensure proper control of Myc activity in normal cells, including regulation of Myc protein turnover, through the ubiquitin-proteosome pathway [69,70]. Recently, a component of the E3 ubiquitin ligase complex, Skp2, has been reported to mediate Myc turnover *in vivo,* which in turn is linked to Myc transcriptional activity at specific target genes. The regions of Skp2–Myc interaction were mapped to MBII as well as to Myc CTD (Fig. 2) [71,72]. Consistent with its role as a co-factor for Myc, Skp2 induces S-phase entry in a Myc-dependent manner [71]. Since Skp2 is known to have oncogenic properties [73,74] its role in increasing Myc transcriptional activity is not surprising. Although Skp2 is required for the phosphorylation-dependent degradation of various proteins, such as the tumour suppressor FOXO1 [75], there is no evidence suggesting the effect of Skp2 on Myc is regulated at the level of phosphorylation. Phosphorylation-dependent degradation of Myc involves two key residues, T58 and S62 within the MBI. Phosphorylation of S62 is mediated by Ras signalling and is a prerequisite to the phosphorylation of T58 that is performed by glycogen synthase kinase 3 (GSK3) [70,76–78]. Phosphorylation at T58 destabilises Myc protein [70,79]. T58 represents a major hotspot for mutations in Burkitt's and other lymphomas [80]. T58-phosphorylated Myc is specifically bound by the F-box protein, Fbw7, and results in Myc ubiquitination and degradation [80–83]. Fbw7 appears to function as a tumour suppressor gene [84]. Knockdown of Fbw7 increased both the abundance and transactivation activity of endogenous Myc. Surprisingly, one of the Fbw7 isoforms (Fbw7γ) co-localises with Myc in the nucleolus upon proteosome inhibition suggesting that Myc is also regulated in this specialised nuclear compartment for degradation [81]. These data support the idea that Fbw7 and Skp2 have opposing effects on Myc activity by targeting MBI and MBII, respectively. This knowledge can be exploited for therapeutics as targeting MBI might affect Myc stability while targeting MBII would affect Myc transcriptional activity.

3. Myc regulated target genes

To understand the role of Myc in cellular physiology and pathology, it is essential to identify the *bona fide* target genes regulated by Myc. A *bona fide* Myc-target gene is one whose regulatory region is bound, directly or indirectly by Myc, and whose expression is then regulated under an appropriate stimulus. These targets are distinguished from the many regulatory events that occur as a downstream consequence of Myc activity, like cell cycle progression [11]. In the past, many criteria were required to distinguish a gene as a true Myc target gene, but none were absolute. Several cDNA microarray studies have identified many *bona fide* targets and downstream regulated genes, but rarely have the two subsets been clearly distinguished [85–88]. Recently developed techniques have revolutionised this major issue and allowed true target genes to be rapidly identified and profiled *in vivo*, in any given cell or tissue under any given stimulus. One such technique is termed ChIP–chip (or ChIP-on-chip), in which the sensitivity and specificity of chromatin immunoprecipitation (ChIP) is combined with the high throughput capability of microarray (chip) technology.

3.1. Genome-wide Myc binding profile

Knowledge of binding sites found within a genome is essential for understanding the target genes regulated by any transcription factor. With the development of high throughput ChIP technologies, many labs, including our own, have identified the *in vivo* genomic DNA-binding sites of Myc [29]. This is especially important for understanding the mechanisms of Myc role in carcinogenesis. By profiling the many target genes regulated by Myc, the anticipation is that a key subset essential for a particular biological function will be distinguished and the mechanism of co-regulation of this cohort of genes will then be determined. Several novel insights have been forthcoming from these studies on Myc genomic binding in mammals [89–91]. It has been concluded that at sites of transcriptional activation, Myc and Max bind together at high affinity loci containing canonical or non-canonical E-boxes, which are often CpG rich [29,92]. Moreover, these characteristics are evolutionarily conserved from *Drosophila* to mammals [93]. In addition, recent work shows that there is evolutionary conservation of promoter architecture across different species of *Drosophila* containing a single E-box located within the first 100 nucleotides downstream of the transcription start site [94]. These outcomes suggest that a subset of Myc targets share a common and particular mechanism of regulation. By contrast, the repression of the target genes by Myc does not occur at E-boxes, but, rather at proximal promoter regions [95–98]. Interestingly, the Myc–Max interaction is essential for Myc to repress as well as activate gene transcription [91]. A systematic analysis across promoters, employing short oligonucleotide arrays, showed that only 22% of the Myc binding sites are located at the 5′ upstream region of protein-coding genes, while 36% are placed at the 3′ end of well-characterised genes that are associated with non-coding RNAs [99]. In light of this study it is very interesting that two recent papers describe the relationship between Myc and non-coding microRNAs (miRNAs). The first shows that Myc regulates a cluster of miRNAs in chromosome 13 that influences the major cell cycle regulator, E2F1 [100]. The authors proposed that Myc directly binds to this miRNA cluster to negatively regulate E2F1, thereby dampening the runaway effect of Myc inducing E2F1 transcription. It will be interesting to determine the function of this miRNA cluster in tumour cells. The second study shows that another cluster in chromosome 13, which is often overexpressed in lymphomas, strongly cooperates with Myc in lymphomagenesis by inhibiting Myc apoptotic capability [101]. The identification of miRNAs is in its infancy and the identification of transcriptionally active sites by ChIP–chip will greatly aid in delineating oncogenically active miRNAs.

One of the most profound results emanating from the genome-wide localisation studies is that Myc binds to an enormous number of target genes, compared with other transcription factors such as p53 and Sp1 [99]. Moreover, the genomic binding sites are associated with genes whose products are engaged in a wide-range of biological processes. In addition, it is curious that not all Myc bound targets are regulated at the level of expression [89].These observations have been gathered from a limited series of experiments conducted in a variety of labs. Clearly, it will be critical to systematically evaluate the nature of Myc target genes and their regulation under a multiplicity of physiological and pathological conditions. Importantly, these experiments are now feasible thanks to the advances in ChIP–chip technologies. To take advantage of these results, it will be necessary to identify Myc interactors that cooperate in the regulation of genes implicated in Myc related phenotypes. The target genes identified to date are compiled in a well-annotated database http://www.myc-cancer-gene.org/site/mycTargetDB.asp [11,102,103]. Due to space constraints, we will only discuss specific *bona fide* targets in the context of understanding their mechanism of regulation by Myc.

3.2. Mechanisms of Myc dependent transcriptional activation

Activation of target genes by Myc involves at least two regulatory steps; chromatin remodeling and promoter clearance. Mechanistic analyses of Myc-induced genes have clearly shown that Myc participates in chro-

matin remodeling when recruited to promoter regions. For example, the activation of the normally silent telomerase reverse transcriptase gene (TERT) by oncogenic Myc in exponentially growing fibroblasts requires TRRAP recruitment and is accompanied by both H3 and H4 acetylation [104]. Despite the profound role of this target gene in oncogenesis, transformation of primary cells is thought to require the regulation of additional targets by oncogenic Myc, as TERT overexpression alone cannot replace Myc in Rat fibroblast transformation assays [105] and mice lacking telomerase RNA can still be transformed by Myc and H-Ras [106]. Many cell cycle components are also regulated by Myc in a MBII/TRRAP-dependent manner [107]. Myc stimulates expression of the Cyclin D2 and Cdk4 genes, leading to sequestration of the cell cycle inhibitor p27 in CyclinD2/Cdk4 complexes [108–110]. Remarkably, p27 is then degraded by the protein product of two Myc activated genes, Cul1 and Cks [109,111,112]. Evidence suggests that the interaction of Myc-TRRAP to recruit HAT activity to target gene promoters such as TERT and Cyclin D2, is important for Myc transformation and may be dispensable for Myc to drive apoptosis [42,44]. For example, MEFs derived from the Cyclin D1/D2 double knock-out mice showed impaired proliferation in response to ectopic Myc expression, but the ability of Myc to potentiate apoptosis remained intact [113].

In addition to directly regulating chromatin remodeling through the recruitment of HATs, Myc also induces target genes involved in the regulation of chromatin dynamics. Recently, two such target genes have been described. MT-MC1 encodes a nuclear protein with homology to certain DNA helicases, and HMG-I encodes one of the high-mobility group proteins. Interestingly, ectopic expression of these proteins in Myc knock-out rat fibroblasts was shown to reconstitute many Myc phenotypes, such as rescuing the parental cell morphology, correction of the slow growth rate, cell size, genomic instability, clonogenicity, tumourigenicity and the regulation of a subset of Myc target genes [114]. Even though these two genes sometimes differ in the reconstitution of Myc functions, significant overlap exists between them. Remarkably, these chromatin remodeling proteins, either individually or in combination, were not able to complement the ability of Myc to potentiate apoptosis following serum withdrawal. These studies leave the door open to a selective investigation of additional fundamental Myc targets, which can restore, for example, the Myc apoptotic function.

Myc has been shown to participate in the transcription regulatory step associated with promoter clearance of RNA polymerase II (RNAP II) [115]. Myc was previously shown to recruit components of the elongation factor P-TEFb (positive transcription elongation factor b) at the *cad* promoter [115,116]. Inhibition of the P-TEFb complex blocked the effects of Myc in transformed cells [117]. Thus, the rate-limiting step of transcriptional activation of specific target genes may be Myc-mediated recruitment of P-TEFb, which then allows phosphorylation of RNAP II and release of a suspended transcription complex. It will be interesting to determine whether a cohort of genes, like *cad*, is similarly regulated by Myc and involved in the control of a common biological activity.

An attractive new development in understanding Myc function as a regulator of gene transcription is the appreciation that Myc not only regulates RNAP II regulated genes, but also affects genes transcribed by RNAP I and III. In fact, the various components of the ribosomal machinery are synthesised by all three RNA polymerases, RNAP I, II and III (Fig. 3). This is consistent with the regulation of ribosomal targets as a common feature of many expression microarray analyses to identify Myc target genes. Ribosomal biogenesis is a fundamental cellular process that takes place in the nucleolus and is essential for ribosome assembly and protein synthesis. Remarkably, the nucleoli are enlarged in cancer cells and several ribosomal proteins are overexpressed in tumours, suggesting a correlation between the deregulation of protein biosynthesis and cancer, leading to an opportunity for the development of innovative therapeutics targeting the translation machine [26]. Mechanistically, Myc binds to TFIIIB, a component of RNAP III machinery, and stimulates RNAP III transcriptional regulation of the 5S rRNA gene [118,119]. In the nucleoli, Myc directly regulates RNAP I transcription by interacting with SL1 (TIF-IB), an essential complex composed of the TATA binding protein (TBP) and three RNAP I-specific TBP-associated factors (TAFs) [120,121]. Additionally, ChIP experiments show that Myc and Max bind at non-canonical E-box sequences located within ribosomal DNA (rDNA) promoters. This association is followed by recruitment of TRRAP, enhanced histone acetylation, recruitment of RNAP I, and activation of rDNA transcription [122–124]. Taken together, regulation of all three RNA polymerases suggest that Myc plays a key role within the cell to produce molecules implicated in ribosome biogenesis (Fig. 3).

3.3. Mechanisms of Myc dependent transcriptional repression

The molecular mechanism of Myc role in repression of gene transcription remains less well characterised than its role in activation, yet all indications suggest repression is as important as activation for Myc function. For example, structure–function analyses have linked transformation to repression. Moreover, microarray and ChIP–chip analyses show that Myc activates and represses gene expression in similar proportions.

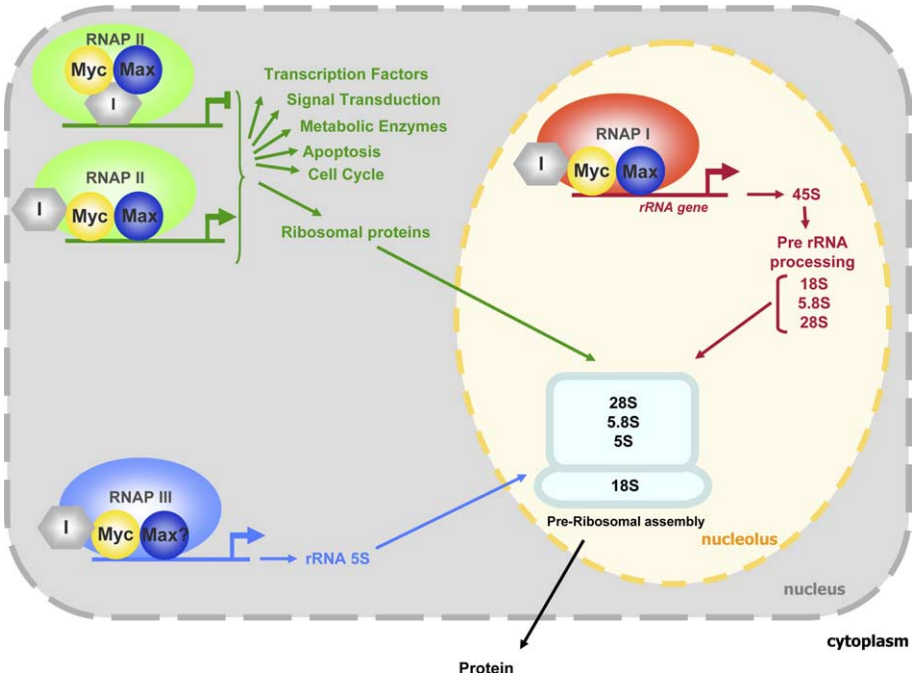


Fig. 3. Myc regulates transcription directed by the three RNA polymerases. Myc, with its partner Max, binds DNA and either activates or represses transcription of numerous target genes of enormous diversity, which are regulated by RNA polymerase II (RNAP II) and implicated in a multiplicity of cellular functions. In the nucleolus, together with Max as a partner, Myc activates the transcription of the rRNA genes, regulated by the RNA polymerase I complex (RNAP I). As well, Myc activates rRNA 5S transcription by interacting with TFIIIB, a component of the RNA polymerase III complex (RNAP III). The collective regulation of these three transcriptional functions by Myc suggests that Myc plays a key role within the cell to produce molecules implicated in ribosome biogenesis. RNAP I, II and III exist as a part of multiprotein complexes: other components are not shown in this simplified figure. The hexagon symbolizes various Myc-protein interactors (I) that contribute to gene regulation. (Figure adapted from Oskarsson and Trumpp [199]).

The target genes repressed by Myc fall into several functional categories. At a molecular level the best characterised targets includes those whose products inhibit cell proliferation (e.g. p15, p21, gadd45). Characterising the mechanism of Myc repression of these and other repressed target genes has shown that Myc does not appear to directly associate with the regulatory regions of repressed targets, but rather is recruited to core promoters through protein–protein interactions. Myc interacts with activating transcription factors, such as TFII-I, NF-Y and Miz-1 [95,125,126]. For example, transcriptional activation by Miz1 is abolished with Myc binding, and the Myc–Miz-1 complex acts as a transcriptional repressor; in part due to competition between p300 and Myc for binding to Miz1 [98]. It was recently shown that Myc represses transcription of p21Cip through recruitment of the DNA methyltransferase corepressor Dnmt3a. Myc and Dnmt3a form a ternary complex with Miz-1 to corepress p21Cip [127]. The precise role of Miz-1, as well as other factors including YY1 [128,129], NF-Y [95,130] and TFII-I [126] in Myc repression remains unclear and will require additional analyses.

Like Myc activation of gene transcription, Myc repression can occur through multiple mechanisms. Myc repression was thought to be dependent upon an initiator (Inr) region within the promoter, however, the presence of an Inr is not essential as genes lacking an Inr, such as gadd45 and PDGFB, are repressed by Myc [95,131]. Myc has been shown to repress gadd45 by a post-RNA polymerase II recruitment mechanism [132]. Interestingly, promoter binding and repression of PDGFBR are separable activities, since mutant Myc proteins that are unable to repress PDGFBR gene expression, still bind to the promoter *in vivo* [130].

Currently a large number of *in vivo* Myc activated or repressed target genes have been compiled from studies in different types of cancer cells. However, many questions remain. For example, which target genes are regulated by Myc and critical for the carcinogenic process? This issue will be essential to further study the subset of Myc regulated genes that directly or indirectly contribute to the tumour formation. Through the characterisation of these target genes, new possibilities will be addressed in the diagnostic field as well as in the development of novel anticancer therapeutics to target the oncogenic activity of Myc.

4. Myc as a target in cancer therapy

Developing therapeutics to inhibit oncogenic Myc would have enormous impact on the treatment of a

wide-range of human cancers. Transgenic mouse models provide a glimpse of how profoundly effective blocking Myc can be as an anti-cancer target *in vivo*. A series of elegant experiments using inducible Myc in hematopoetic cells [133], mammary gland [134], liver [135], skin [135,136] and pancreatic islets [137] have demonstrated that induction of oncogenic Myc leads to full-blown malignancies, while blocking Myc activation in most cases results in tumour regression [138]. This strongly supports the notion that targeting Myc in tumours represents a valid therapeutic approach. Caution should be exercised as not all tumours regressed upon withdrawal of Myc and a small proportion proceeded to hematological malignancies [133], while half of Myc-induced mammary carcinomas acquired mutations in K-Ras or H-Ras, thus rendering them independent of Myc status [134]. It is likely that anti-Myc agents would have to be coupled with therapies targeting Ras or other oncogenic pathways.

Many strategies are under development to target or exploit oncogenic Myc in tumour cells and eradicate the malignant cellular clones (Fig. 4). One approach is to disrupt Myc expression by targeting regulatory steps ranging from transcription to translation. Another strategy is to block Myc function by inhibiting critical protein–protein interactions that are essential for Myc to regulate gene transcription, such as heterodimerisation with Max. The working assumption is that targeting the Myc regulatory network will trigger the lethality of the tumour cell without causing irreversible damage to neighbouring normal cells. Such a tumour-specific effect is evident with other inhibitors that target universally expressed oncogenes. For example, antisense *bcl-2* will sensitise tumour cells to undergo apoptosis in response to low-dose chemotherapy, but non-transformed cells are spared [139–142]. In this case, the tumour cell has become dependent upon the deregulated signalling pathway and even marginal down-regulation of this lifeline renders it susceptible to extinction. Should this assumption be incorrect, Myc inhibitors will have to be targeted by tumour-specific delivery mechanisms. Newer strategies aim to achieve a high therapeutic index by targeting tumour-specific Myc-protein interactions and/or gene regulatory functions. Yet others aim to exploit the presence of oncogenic Myc expression in tumour cells to specifically trigger a suicide response exclusively in these transformed cells. Both ongoing and emerging strategies will be further discussed below.

4.1. Targeting Myc expression

One of the first successful applications of antisense technology targeted Myc expression, and through advancements over the last 15 years, this approach remains at the forefront of anti-Myc therapeutics [8,143]. Antisense oligonucleotides (ASOs) are short single-stranded DNA molecules that specifically target, hybridize and inhibit the mRNA of a selected gene (Fig. 4).

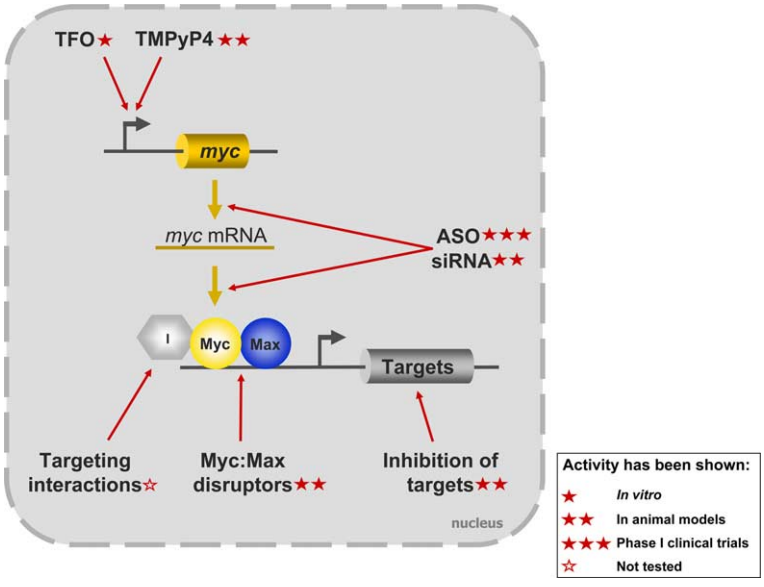


Fig. 4. Strategies for targeting oncogenic Myc in cancer. Therapeutic agents are coded according to their current therapeutic stage of development. Agents that have demonstrated efficacy *in vitro* are indicated with one star. These include inhibitors that block Myc expression, such as triple helix forming oligonucleotides (TFO). Double star agents have been tested in animal tumour models or xenograft models for their efficacy to inhibit Myc. This category includes disruptors of Myc–Max interaction and agents that block Myc target gene function, such as cationic porphyrins (TMPyP4) and small interfering RNAs (siRNA). Some of the antisense oligonucleotides (ASO) that have successfully completed Phase I clinical trials, and are at advanced stage of drug development are indicated by three stars. Finally, a clear star indicates additional potential targets for the therapeutic interference of Myc–protein interactors (I).

The DNA–RNA duplex then recruits RNase H endonuclease to cleave the RNA strand in the duplex [144], leaving the antisense DNA intact to hybridize to additional mRNAs of the target gene [145,146]. Antisense technology has evolved as a viable therapeutic alternative by increasing the functional stability and permeability of the ASOs. This has been largely achieved by replacing the phosphodiester backbone with a nuclease resistant phosphorothioate linkage (PS ASOs) [147,148].

Myc ASOs have advanced successfully through the many stages of preclinical evaluation and have demonstrated anti-cancer potential. Evaluation *in vitro* has shown that downregulating Myc expression by PS ASOs reduces leukaemic cell proliferation, induces differentiation and inhibits G1/S progression [149,150]. Moreover, exposure of MCF-7 breast cancer cells to Myc ASOs induced growth arrest [151]. In mouse models, treatment with Myc PS ASOs resulted in the delay or prevention of Burkitt's lymphoma [152–154] and Myc ASOs enhanced the efficacy of cisplatin to target melanoma both *in vitro* and *in vivo* [155]. Cisplatin resistance can be overcome by inhibiting Myc using AVI-4126, a phosphorodiamidate morpholino oligomer (PMO), in the Lewis lung carcinoma model [156]. AVI-4126 also inhibited growth of a murine prostate cancer xenograft by inducing growth inhibition and apoptosis *in vivo* [157]. These results have led to Phase I clinical trials, which show that intravenous administration of the morpholino oligomer was not accompanied by toxicity or serious adverse events, and importantly, the bioavailability was measurable in malignant tumours. The number of patients studied remains modest, yet this data supports the use of the AVI-4126 PMO as a potential therapeutic for cancer treatment [158]. It will be interesting to monitor the efficacy of this agent as it advances through to Phase II/III clinical trials.

Several additional Myc ASOs show promise for future application as anti-cancer agents. Administration of INX-6295, a 16-mer Myc PS ASO encapsulated in lipid particles, shows antitumour efficacy against a human melanoma xenograft. When administered with cisplatin, INX-6295 produced complete tumour regression in 30% of treated mice [159]. Furthermore, the combined application of *bcl-2* ASO/cisplatin/INX-6295 in mice harbouring human melanoma xenografts overexpressing either *bcl-2* or *c-myc* oncogenes resulted in effective antitumour therapy [160]. Yet another agent, the novel psoralen- or acridine-modified, clamp-forming ASOs, can downregulate Myc expression and synergise with cisplatin to inhibit melanoma cell proliferation and tumour progression [161,162]. The peptide nucleic acid (PNA) ASO also shows promise. PNA is a synthetic DNA in which the sugar-phosphate backbone is replaced with a polyamide-(2-aminoethyl) glycine-skeleton. This modification provides PNAs with a long half-life and enables PNAs to specifically hybridise to DNA and/or RNA in a complementary manner, forming a strong and effective duplex that can inhibit transcription and translation of the target gene [163]. Myc expression was rapidly downregulated in Burkitt's Lymphoma, using a 17-mer anti-*c-myc* PNA covalently linked to a nuclear localisation signal (NLS) (PNA-myc_{wt}-NLS) [164]. However, because this PNA blocked Myc in both transformed and non-transformed cells, further innovative refinements of this strategy are under development to increase tumour specificity [165]. Clearly, ASO technology targeting Myc expression is well advanced and has enormous promise for future application to patient care. Maximal efficacy will depend on achieving synergy with conventional chemotherapies and/or novel molecular anti-cancer agents [166,167].

RNA interference (RNAi) is a modern and popular approach to knockdown gene expression that has potential for drug development. RNAi is a mechanism for silencing gene expression through targeting double-stranded RNA to mRNA resulting in degradation of the targeted mRNA (Fig. 4). In mammalian cells, long double-strand RNAs are cleaved into small interfering RNAs (siRNAs) that, through incorporation into the silencing complex, direct target recognition. RNAi has been touted as the next major tool in targeted cancer therapy, because of its impressive specificity and efficacy. Compared to antisense approaches, siRNAs are 1000-fold more active [168] and a plethora of data shows the efficacy of siRNAs in cell culture. For example, Myc siRNA effectively inhibited Myc protein levels in MCF7 cells [169]. It appears that siRNA can be effective *in vivo* [170], but this has not yet been thoroughly explored. The primary limitations of siRNA *in vivo* remain its stability and delivery. siRNA is quickly degraded in plasma, so the duplex will need to be chemically modified for use as a drug. Perhaps strategies used to modify ASOs to increase stability can be exploited for siRNA. Recently the issue of delivery was addressed by fusing multiple siRNAs, targeting Myc, Hdm2 and VEGF mRNAs, to a positively charged protein that was covalently linked to a specific antibody. This ensured that only melanoma cells ectopically expressing that particular ligand were growth inhibited *in vivo* [171]. With this novel approach it may be possible to deliver such a lethal siRNA cocktail through a Myc-induced cell surface molecule to achieve tumour-specific targeting. Whether siRNAs target mutated oncogenes like Ras [172] and/or deregulated oncogenes like Myc [169], these new tools are likely to make a firm and lasting entrance into the arsenal of therapeutics in the fight against cancer.

Another series of agents have been designed to target *myc* expression at the transcriptional level. One promising agent includes the triple helix forming oligonucleotides (TFOs) which bind to double-stand purine-rich DNA within promoter regions and block transcription factor binding (Fig. 4). Phosphothioate stabilised TFOs

directed against the promoter region of c-*myc* have been shown to successfully inhibit Myc expression in several leukaemia and lymphoma cell lines leading to cellular growth arrest and apoptosis [173,174]. TFOs conjugated to the DNA-intercalating agent daunomycin specifically downregulated endogenous Myc in prostate and breast cancer cell lines [175]. Cationic porphyrin TMPyP4, which inhibits Myc transcription by blocking G quadruplexes, a DNA structure formed from G rich single-strand DNA during active transcription is another approach to target Myc [176] (Fig. 4). TMPyP4 inhibited the *in vitro* transcription of *myc* and decreased tumour growth rates in xenograft models [176]. The original use of cationic porphyrin was to block telomerase by stabilising the telomeric G rich single stand DNA overhangs into G quadruplexes [176]. New Myc-specific porphyrin analogues show promising results, suggesting that this approach has merit for further development [177,178]. Introduction of G quadruplex-forming oligonucleotides into Burkitt's lymphoma cells resulted in growth inhibition by sequestering the factors that normally bind to the native G quadruplexes formed at the Myc promoter [179]. It will be interesting to monitor the development of these various agents that target *myc* transcription for their ultimate use as therapeutics.

4.2. Targeting Myc–Max interaction

Breaking the Myc–Max bond would clearly destroy oncogenic Myc function and several strategies to dissolve this partnership are well underway (Fig. 4). Many of the approaches under development have been advanced because of our knowledge of the detailed structural biology of this protein–protein interaction and the essential residues involved. Molecular modeling and mutagenesis have been used to identify specific amino acids that alter the specificity of dimerisation [180]. Four amino acids within the LZ of Myc were sufficient for homodimerisation, as well as heterodimerisation with Myc and Max. Furthermore, a mutant protein termed Omomyc, which interferes with Myc binding to E-box elements, was able to inhibit colony formation in NIH3T3 cells [181]. Introduction of Omomyc in a Myc-induced skin tumourigenesis model in mice [182] inhibited Myc-induced papillomatosis, as well as restored the normal keratinocyte differentiation program and skin architecture, both of which are otherwise disrupted by Myc activation [183]. Importantly, the positive outcome of Omomyc expression was completely dependent on the presence of an oncogenic Myc and therefore may be of therapeutic value for the specific targeting of Myc-deregulated cells without affecting the surrounding normal cells.

Knowing the structural domains required for the Myc–Max interaction inspired a dominant-negative approach to disrupt this linkage. Myc mutants expressing only the b-HLH-LZ or HLH-LZ domain rapidly induced apoptosis in 3T3-L1 mouse fibroblasts [184]. This study provided the rationale to devise additional strategies to directly exploit Myc heterodimerisation for the design of novel therapeutics. One of the major challenges with this approach lies in achieving efficient delivery of the drug to the nucleus of tumour cells *in vivo*. To this end, a mutant peptide derived from the helix-1 region of Myc was linked to an internalisation sequence [185]. The fusion peptide interfered with the transcriptional activity of Myc leading to the inhibition of MCF-7 cell growth. The stability and activity of this peptidomimetic molecule was increased [186,187] and a variant of the original peptidomimetic has been synthesised and tested in mouse models. Interestingly, analysis of inhibitor interactions of Myc–Max shows the active molecules act through key basic residues at the outer surface of the Myc–Max heterodimer, potentially by binding or interfering with another interactor. The novel peptides were capable of reaching high concentrations in mouse organs and were effective at inhibiting growth of a colon cancer cell line. This peptide is 10-fold larger than traditional small molecules, which may contribute to its highly selective interference with Myc-specific protein–protein interactions [186]. It would be of great interest to test which specific Myc–protein interaction is inhibited *in vivo* and whether knockdown of this co-factor can directly trigger an anti-tumour effect in animals.

Recently, two groups have shown that inhibitors blocking the Myc–Max interaction can be isolated using a high throughput screen. In one study, inhibitors were identified by fluorescence resonance energy transfer in high-throughput screens of peptidomimetic libraries, then confirmed by enzyme-linked immunosorbent assay and electrophoretic mobility shift assay. The antagonists interfered with Myc-induced and Jun-induced oncogenic transformation suggesting the inhibitor may also target the LZ of the Jun oncoprotein [188]. The exact molecular mechanism of inhibition and the utility of these agents in the control of carcinogenesis is the focus of future analysis. In the second study, the yeast two-hybrid assay was used to screen a library of 10 000 compounds to identify those able to disrupt the interaction of the b-HLH-LZ regions of Myc and Max, yet show no general toxicity to the yeast [189]. Several compounds were identified and their ability to inhibit the Myc–Max interaction confirmed by an *in vitro* association assay. Using a reporter assay, the compounds were shown to inhibit Myc transcriptional activity and proved to inhibit growth of Myc transformed rat fibroblasts, but not Myc-null cells. Finally, incubation of transformed cells with the compounds for three days, prior to their injection into nude mice, inhibited tumour growth *in vivo*. Although the compounds used by both groups require relatively high concentrations, these promising studies

provide a platform for future development of more effective small molecule inhibitors of Myc–Max dimerisation.

5. Conclusions and perspectives

Targeting Myc at the level of expression and/or function is an effective approach to eliminate this potent oncoprotein. The agents described in this review successfully block Myc and have passed the first hurdle in targeted drug design. The issues that remain to be resolved include, but are not limited to, evaluating and improving sensitivity, specificity, delivery and efficacy as a single agent and in combination with other anti-neoplastic therapies. The agents that have advanced beyond Phase I clinical trials show enormous promise. It will be fascinating to monitor their progress and learn whether targeting Myc will have the expected impact at the level of patient care.

There are several additional approaches aimed at targeting Myc activity that are in the earliest stages of development. Based on the success of targeting Myc–Max interaction, there is significant interest in targeting other key Myc–protein partnerships, such as Myc-TRRAP. This field of study covers a broad spectrum of experimental analyses, from first identifying the interactor, determining whether it plays an important functional role in Myc transformation, through to mapping the precise points of interaction. Many Myc–protein interactions highlighted here have been described only in recent years and some of them may warrant inhibitor development.

A second major area of fundamental research that has the potential to impact therapeutic design aims at identifying and understanding Myc target genes and the biological pathways they regulate. As a multifunctional master regulator, Myc induces several genes which, in turn, play a critical role in transformation and are now being targeted as potential anti-cancer therapeutics. For example, haplo insufficiency for a Myc target gene, *odc* reduces skin tumourigenesis in mice [190] and a specific inhibitor of ODC, 2-difluoromethylornitive, can block Myc-induced oncogenesis [191]. Clearly, a greater understanding of the Myc transformation program will result in additional opportunities to target Myc function in tumour cells.

Another novel approach does not target Myc directly, but instead aims to exploit the ability of oncogenic Myc to potentiate apoptosis. A cooperating lesion, such as Bcl-2 activation, often inhibits Myc potentiation of apoptosis and collaborates with Myc to drive transformation. Blocking these anti-apoptotic molecules may release the ability of Myc to sensitise tumour cells to undergo apoptosis, elevate the therapeutic index and achieve tumour cell death. There is enormous effort to understand the nature of the genetic abnormalities associated with cancers that can block Myc-potentiation of apoptosis and contribute to transformation. With this knowledge, the innovative approach of exploiting deregulated Myc can be further advanced.

Myc deregulation is often associated with aggressive disease of poor prognosis, which augments the urgency for novel therapeutics targeting this potent oncoprotein. Many such agents are well along the drug development pipeline. Moreover, fundamental research instructs us daily of new opportunities to effectively target Myc expression and function to block malignant transformation. Further growth in the area of anti-Myc therapeutics is warranted and anticipated.

Conflict of interest statement

None declared.

Acknowledgements

We apologize to all colleagues for the work we could not cite due to space constraints. We thank the Penn Lab for constructive comments and critical reading of this manuscript and P. De Luca for her excellent editorial work. S.K. is a research fellow of the Terry Fox Foundation through an award from the National Cancer Institute of Canada (NCIC), D.B.-L. holds a postdoctoral fellowship from the Canadian Institute of Health Research (CIHR). This work was supported by grants from the NCIC with funds from the Canadian Cancer Society, US Department of Defense (DOD), and CIHR (L.Z.P.).

References

1. Emens LA. Trastuzumab: targeted therapy for the management of HER-2/neu-overexpressing metastatic breast cancer. *Am J Ther* 2005, **12**(3), 243–253.
2. Krause DS, Van Etten RA. Tyrosine kinases as targets for cancer therapy. *N Engl J Med* 2005, **353**(2), 172–187.
3. Ocana A, Rodriguez CA, Cruz JJ. Integrating trastuzumab in the treatment of breast cancer. Current status and future trends. *Clin Transl Oncol* 2005, **7**(3), 99–100.
4. Sawyers C. Targeted cancer therapy. *Nature* 2004, **432**(7015), 294–297.
5. Dang CV. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol Cell Biol* 1999, **19**(1), 1–11.
6. Prochownik EV. c-Myc as a therapeutic target in cancer. *Expert Rev Anticancer Ther* 2004, **4**(2), 289–302.
7. Eilers M, Picard D, Yamamoto KR, *et al*. Chimaeras of myc oncoprotein and steroid receptors cause hormone-dependent transformation of cells. *Nature* 1989, **340**(6228), 66–68.
8. Heikkila R, Schwab G, Wickstrom E, *et al*. A c-myc antisense oligodeoxynucleotide inhibits entry into S phase but not progress from G0 to G1. *Nature* 1987, **328**(6129), 445–449.

9. Roussel MF. Key effectors of signal transduction and G1 progression. *Adv Cancer Res* 1998, **74**, 1–24.
10. Adachi S, Obaya AJ, Han Z, *et al*. c-Myc is necessary for DNA damage-induced apoptosis in the G (2) phase of the cell cycle. *Mol Cell Biol* 2001, **21**(15), 4929–4937.
11. Oster SK, Ho CS, Soucie EL, *et al*. The myc oncogene: MarvelouslY Complex. *Adv Cancer Res* 2002, **84**, 81–154.
12. Cole MD. The myc oncogene: its role in transformation and differentiation. *Annu Rev Genet* 1986, **20**, 361–384.
13. Boxer LM, Dang CV. Translocations involving c-myc and c-myc function. *Oncogene* 2001, **20**(40), 5595–5610.
14. Nesbit CE, Tersak JM, Prochownik EV. MYC oncogenes and human neoplastic disease. *Oncogene* 1999, **18**(19), 3004–3016.
15. Schwab M. MYCN in neuronal tumours. *Cancer Lett* 2004, **204**(2), 179–187.
16. Thomas WD, Raif A, Hansford L, *et al*. N-myc transcription molecule and oncoprotein. *Int J Biochem Cell Biol* 2004, **36**(5), 771–775.
17. Wu R, Lin L, Beer DG, *et al*. Amplification and overexpression of the L-MYC proto-oncogene in ovarian carcinomas. *Am J Pathol* 2003, **162**(5), 1603–1610.
18. Blakely CM, Sintasath L, D'Cruz CM, *et al*. Developmental stage determines the effects of MYC in the mammary epithelium. *Development* 2005, **132**(5), 1147–1160.
19. Donaldson TD, Duronio RJ. Cancer cell biology: Myc wins the competition. *Curr Biol* 2004, **14**(11), R425–R427.
20. Flores I, Murphy DJ, Swigart LB, *et al*. Defining the temporal requirements for Myc in the progression and maintenance of skin neoplasia. *Oncogene* 2004, **23**(35), 5923–5930.
21. Hoffman B, Amanullah A, Shafarenko M, *et al*. The proto-oncogene c-myc in hematopoietic development and leukemogenesis. *Oncogene* 2002, **21**(21), 3414–3421.
22. Mai S, Mushinski JF. c-Myc-induced genomic instability. *J Environ Pathol Toxicol Oncol* 2003, **22**(3), 179–199.
23. Nilsson JA, Cleveland JL. Myc pathways provoking cell suicide and cancer. *Oncogene* 2003, **22**(56), 9007–9021.
24. Pelengaris S, Khan M, Evan G. c-MYC- more than just a matter of life and death. *Nat Rev Cancer* 2002, **2**(10), 764–776.
25. Ruddell A, Mezquita P, Brandvold KA, *et al*. B lymphocyte-specific c-Myc expression stimulates early and functional expansion of the vasculature and lymphatics during lymphomagenesis. *Am J Pathol* 2003, **163**(6), 2233–2245.
26. Schmidt EV. The role of c-myc in regulation of translation initiation. *Oncogene* 2004, **23**(18), 3217–3221.
27. Secombe J, Pierce SB, Eisenman RN. Myc – a weapon of mass destruction. *Cell* 2004, **117**(2), 153–156.
28. Soucek L, Evan G. Myc – is this the oncogene from Hell? *Cancer Cell* 2002, **1**(5), 406–408.
29. Patel JH, Loboda AP, Showe MK, *et al*. Analysis of genomic targets reveals complex functions of MYC. *Nat Rev Cancer* 2004, **4**(7), 562–568.
30. Luscher B. Function and regulation of the transcription factors of the Myc/Max/Mad network. *Gene* 2001, **277**(1–2), 1–14.
31. Dang CV, Dolde C, Gillison ML, *et al*. Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. *Proc Natl Acad Sci USA* 1992, **89**(2), 599–602.
32. Grandori C, Cowley SM, James LP, *et al*. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu Rev Cell Dev Biol* 2000, **16**, 653–699.
33. Conzen SD, Gottlob K, Kandel ES, *et al*. Induction of cell cycle progression and acceleration of apoptosis are two separable functions of c-Myc- transrepression correlates with acceleration of apoptosis. *Mol Cell Biol* 2000, **20**(16), 6008–6018.
34. Herbst A, Hemann MT, Tworkowski KA, *et al*. A conserved element in Myc that negatively regulates its proapoptotic activity. *EMBO Rep* 2005, **6**(2), 177–183.
35. Fladvad M, Zhou K, Moshref A, *et al*. N and C-terminal subregions in the c-Myc transactivation region and their joint role in creating versatility in folding and binding. *J Mol Biol* 2005, **346**(1), 175–189.
36. McEwan IJ, Dahlman-Wright K, Ford J, *et al*. Functional interaction of the c-Myc transactivation domain with the TATA binding protein: evidence for an induced fit model of transactivation domain folding. *Biochemistry* 1996, **35**(29), 9584–9593.
37. Blackwood EM, Eisenman RN. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* 1991, **251**(4998), 1211–1217.
38. Nair SK, Burley SK. X-ray structures of Myc–Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell* 2003, **112**(2), 193–205.
39. Brownlie P, Ceska T, Lamers M, *et al*. The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. *Structure* 1997, **5**(4), 509–520.
40. Amati B, Frank SR, Donjerkovic D, *et al*. Function of the c-Myc oncoprotein in chromatin remodeling and transcription. *Biochim Biophys Acta* 2001, **1471**(3), M135–M145.
41. McMahon SB, Van Buskirk HA, Dugan KA, *et al*. The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. *Cell* 1998, **94**(3), 363–374.
42. Park J, Kunjibettu S, McMahon SB, *et al*. The ATM-related domain of TRRAP is required for histone acetyltransferase recruitment and Myc-dependent oncogenesis. *Genes Dev* 2001, **15**(13), 1619–1624.
43. Grant PA, Schieltz D, Pray-Grant MG, *et al*. The ATM-related cofactor Tra1 is a component of the purified SAGA complex. *Mol Cell* 1998, **2**(6), 863–867.
44. McMahon SB, Wood MA, Cole MD. The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc. *Mol Cell Biol* 2000, **20**(2), 556–562.
45. Fuchs M, Gerber J, Drapkin R, *et al*. The p400 complex is an essential E1A transformation target. *Cell* 2001, **106**(3), 297–307.
46. Vervoorts J, Luscher-Firzlaff JM, Rottmann S, *et al*. Stimulation of c-MYC transcriptional activity and acetylation by recruitment of the cofactor CBP. *EMBO Rep* 2003, **4**(5), 484–490.
47. Wood MA, McMahon SB, Cole MD. An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc. *Mol Cell* 2000, **5**(2), 321–330.
48. Cheng SW, Davies KP, Yung E, *et al*. c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function. *Nat Genet* 1999, **22**(1), 102–105.
49. Sevenet N, Lellouch-Tubiana A, Schofield D, *et al*. Spectrum of hSNF5/INI1 somatic mutations in human cancer and genotype–phenotype correlations. *Hum Mol Genet* 1999, **8**(13), 2359–2368.
50. Sevenet N, Sheridan E, Amram D, *et al*. Constitutional mutations of the hSNF5/INI1 gene predispose to a variety of cancers. *Am J Hum Genet* 1999, **65**(5), 1342–1348.
51. Versteege I, Sevenet N, Lange J, *et al*. Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* 1998, **394**(6689), 203–206.
52. Park J, Wood MA, Cole MD. BAF53 forms distinct nuclear complexes and functions as a critical c-Myc-interacting nuclear cofactor for oncogenic transformation. *Mol Cell Biol* 2002, **22**(5), 1307–1316.
53. Sakamuro D, Elliott KJ, Wechsler-Reya R, *et al*. BIN1 is a novel MYC-interacting protein with features of a tumour suppressor. *Nat Genet* 1996, **14**(1), 69–77.
54. Sakamuro D, Prendergast GC. New Myc-interacting proteins: a second Myc network emerges. *Oncogene* 1999, **18**(19), 2942–2954.
55. Pineda-Lucena A, Ho CS, Mao DY, *et al*. A structure-based model of the c-Myc/Bin1 protein interaction shows alternative

splicing of Bin1 and c-Myc phosphorylation are key binding determinants. *J Mol Biol* 2005, **351**(1), 182–194.
56. Hirst M, Ho C, Sabourin L, *et al*. A two-hybrid system for transactivator bait proteins. *Proc Natl Acad Sci USA* 2001, **98**(15), 8726–8731.
57. Huang A, Ho CS, Ponzielli R, *et al*. Identification of a novel c-Myc protein interactor, JPO2, with transforming activity in medulloblastoma cells. *Cancer Res* 2005, **65**(13), 5607–5619.
58. Osthus RC, Karim B, Prescott JE, *et al*. The Myc target gene JPO1/CDCA7 is frequently overexpressed in human tumours and has limited transforming activity in vivo. *Cancer Res* 2005, **65**(13), 5620–5627.
59. Prescott JE, Osthus RC, Lee LA, *et al*. A novel c-Myc-responsive gene, JPO1, participates in neoplastic transformation. *J Biol Chem* 2001, **276**(51), 48276–48284.
60. Broder YC, Katz S, Aronheim A. The ras recruitment system, a novel approach to the study of protein–protein interactions. *Curr Biol* 1998, **8**(20), 1121–1124.
61. Hennemann H, Vassen L, Geisen C, *et al*. Identification of a novel Kruppel-associated box domain protein, Krim-1, that interacts with c-Myc and inhibits its oncogenic activity. *J Biol Chem* 2003, **278**(31), 28799–28811.
62. Datta A, Nag A, Pan W, *et al*. Myc-ARF (alternate reading frame) interaction inhibits the functions of Myc. *J Biol Chem* 2004, **279**(35), 36698–36707.
63. Qi Y, Gregory MA, Li Z, *et al*. p19ARF directly and differentially controls the functions of c-Myc independently of p53. *Nature* 2004, **431**(7009), 712–717.
64. Smith KP, Byron M, O'Connell BC, *et al*. c-Myc localisation within the nucleus: evidence for association with the PML nuclear body. *J Cell Biochem* 2004, **93**(6), 1282–1296.
65. Cairo S, De Falco F, Pizzo M, *et al*. PML interacts with Myc, and Myc target gene expression is altered in PML-null fibroblasts. *Oncogene* 2005, **24**(13), 2195–2203.
66. Ching RW, Dellaire G, Eskiw CH, *et al*. PML bodies: a meeting place for genomic loci? *J Cell Sci* 2005, **118**(Pt 5), 847–854.
67. Yu M, Schreek S, Cerni C, *et al*. PARP-10, a novel Myc-interacting protein with poly (ADP-ribose) polymerase activity, inhibits transformation. *Oncogene* 2005, **24**(12), 1982–1993.
68. Veldman T, Liu X, Yuan H, *et al*. Human papillomavirus E6 and Myc proteins associate in vivo and bind to and cooperatively activate the telomerase reverse transcriptase promoter. *Proc Natl Acad Sci USA* 2003, **100**(14), 8211–8216.
69. Salghetti SE, Caudy AA, Chenoweth JG, *et al*. Regulation of transcriptional activation domain function by ubiquitin. *Science* 2001, **293**(5535), 1651–1653.
70. Salghetti SE, Kim SY, Tansey WP. Destruction of Myc by ubiquitin-mediated proteolysis: cancer-associated and transforming mutations stabilise Myc. *EMBO J* 1999, **18**(3), 717–726.
71. Kim SY, Herbst A, Tworkowski KA, *et al*. Skp2 regulates Myc protein stability and activity. *Mol Cell* 2003, **11**(5), 1177–1188.
72. von der Lehr N, Johansson S, Wu S, *et al*. The F-box protein Skp2 participates in c-Myc proteosomal degradation and acts as a cofactor for c-Myc-regulated transcription. *Mol Cell* 2003, **11**(5), 1189–1200.
73. Gstaiger M, Jordan R, Lim M, *et al*. Skp2 is oncogenic and overexpressed in human cancers. *Proc Natl Acad Sci USA* 2001, **98**(9), 5043–5048.
74. Latres E, Chiarle R, Schulman BA, *et al*. Role of the F-box protein Skp2 in lymphomagenesis. *Proc Natl Acad Sci USA* 2001, **98**(5), 2515–2520.
75. Huang H, Regan KM, Wang F, *et al*. Skp2 inhibits FOXO1 in tumour suppression through ubiquitin-mediated degradation. *Proc Natl Acad Sci USA* 2005, **102**(5), 1649–1654.
76. Gregory MA, Qi Y, Hann SR. Phosphorylation by glycogen synthase kinase-3 controls c-myc proteolysis and subnuclear localisation. *J Biol Chem* 2003, **278**(51), 51606–51612.
77. Lutterbach B, Hann SR. Hierarchical phosphorylation at N-terminal transformation-sensitive sites in c-Myc protein is regulated by mitogens and in mitosis. *Mol Cell Biol* 1994, **14**(8), 5510–5522.
78. Sears RC. The life cycle of C-Myc – from synthesis to degradation. *Cell Cycle* 2004, **3**(9), 1133–1137.
79. Gregory MA, Hann SR. c-Myc proteolysis by the ubiquitin-proteosome pathway: stabilisation of c-Myc in Burkitt's lymphoma cells. *Mol Cell Biol* 2000, **20**(7), 2423–2435.
80. Bhatia K, Huppi K, Spangler G, *et al*. Point mutations in the c-Myc transactivation domain are common in Burkitt's lymphoma and mouse plasmacytomas. *Nat Genet* 1993, **5**(1), 56–61.
81. Welcker M, Orian A, Grim JA, *et al*. A nucleolar isoform of the Fbw7 ubiquitin ligase regulates c-Myc and cell size. *Curr Biol* 2004, **14**(20), 1852–1857.
82. Welcker M, Orian A, Jin J, *et al*. The Fbw7 tumour suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation. *Proc Natl Acad Sci USA* 2004, **101**(24), 9085–9090.
83. Yada M, Hatakeyama S, Kamura T, *et al*. Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. *EMBO J* 2004, **23**(10), 2116–2125.
84. Sears R, Leone G, DeGregori J, *et al*. Ras enhances Myc protein stability. *Mol Cell* 1999, **3**(2), 169–179.
85. Guo QM, Malek RL, Kim S, *et al*. Identification of c-myc responsive genes using rat cDNA microarray. *Cancer Res* 2000, **60**(21), 5922–5928.
86. O'Connell BC, Cheung AF, Simkevich CP, *et al*. A large scale genetic analysis of c-Myc-regulated gene expression patterns. *J Biol Chem* 2003, **278**(14), 12563–12573.
87. Watson JD, Oster SK, Shago M, *et al*. Identifying genes regulated in a Myc-dependent manner. *J Biol Chem* 2002, **277**(40), 36921–36930.
88. Yu Q, He M, Lee NH, *et al*. Identification of Myc-mediated death response pathways by microarray analysis. *J Biol Chem* 2002, **277**(15), 13059–13066.
89. Fernandez PC, Frank SR, Wang L, *et al*. Genomic targets of the human c-Myc protein. *Genes Dev* 2003, **17**(9), 1115–1129.
90. Li Z, Van Calcar S, Qu C, *et al*. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc Natl Acad Sci USA* 2003, **100**(14), 8164–8169.
91. Mao DY, Watson JD, Yan PS, *et al*. Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression. *Curr Biol* 2003, **13**(10), 882–886.
92. Blackwell TK, Huang J, Ma A, *et al*. Binding of myc proteins to canonical and noncanonical DNA sequences. *Mol Cell Biol* 1993, **13**(9), 5216–5224.
93. Orian A, van Steensel B, Delrow J, *et al*. Genomic binding by the *Drosophila* Myc, Max, Mad/Mnt transcription factor network. *Genes Dev* 2003, **17**(9), 1101–1114.
94. Hulf T, Bellosta P, Furrer M, *et al*. Whole-genome analysis reveals a strong positional bias of conserved dMyc-dependent E-boxes. *Mol Cell Biol* 2005, **25**(9), 3401–3410.
95. Izumi H, Molander C, Penn LZ, *et al*. Mechanism for the transcriptional repression by c-Myc on PDGF β-receptor. *J Cell Sci* 2001, **114**(Pt. 8), 1533–1544.
96. Oster SK, Marhin WW, Asker C, *et al*. Myc is an essential negative regulator of platelet-derived growth factor β receptor expression. *Mol Cell Biol* 2000, **20**(18), 6768–6778.
97. Seoane J, Le HV, Massague J. Myc suppression of the p21 (Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage. *Nature* 2002, **419**(6908), 729–734.
98. Staller P, Peukert K, Kiermaier A, *et al*. Repression of p15INK4b expression by Myc through association with Miz-1. *Nat Cell Biol* 2001, **3**(4), 392–399.
99. Cawley S, Bekiranov S, Ng HH, *et al*. Unbiased mapping of transcription factor binding sites along human chromosomes 21

and 22 points to widespread regulation of noncoding RNAs. *Cell* 2004, **116**(4), 499–509.
100. O'Donnell KA, Wentzel EA, Zeller KI, *et al*. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 2005, **435**(7043), 839–843.
101. He L, Thomson JM, Hemann MT, *et al*. A microRNA polycistron as a potential human oncogene. *Nature* 2005, **435**(7043), 828–833.
102. The Myc Target Gene Database. <http://www.myc-cancer-gene.org/site/mycTargetDB.asp>.
103. Zeller KI, Jegga AG, Aronow BJ, *et al*. An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome Biol* 2003, **4**(10), R69.
104. Nikiforov MA, Chandriani S, Park J, *et al*. TRRAP-dependent and TRRAP-independent transcriptional activation by Myc family oncoproteins. *Mol Cell Biol* 2002, **22**(14), 5054–5063.
105. Greenberg RA, O'Hagan RC, Deng H, *et al*. Telomerase reverse transcriptase gene is a direct target of c-Myc but is not functionally equivalent in cellular transformation. *Oncogene* 1999, **18**(5), 1219–1226.
106. Blasco MA, Lee HW, Hande MP, *et al*. Telomere shortening and tumour formation by mouse cells lacking telomerase RNA. *Cell* 1997, **91**(1), 25–34.
107. Bouchard C, Dittrich O, Kiermaier A, *et al*. Regulation of cyclin D2 gene expression by the Myc/Max/Mad network: Myc-dependent TRRAP recruitment and histone acetylation at the cyclin D2 promoter. *Genes Dev* 2001, **15**(16), 2042–2047.
108. Bouchard C, Thieke K, Maier A, *et al*. Direct induction of cyclin D2 by Myc contributes to cell cycle progression and sequestration of p27. *EMBO J* 1999, **18**(19), 5321–5333.
109. Coller HA, Grandori C, Tamayo P, *et al*. Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signalling, and adhesion. *Proc Natl Acad Sci USA* 2000, **97**(7), 3260–3265.
110. Hermeking H, Rago C, Schuhmacher M, *et al*. Identification of CDK4 as a target of c-MYC. *Proc Natl Acad Sci USA* 2000, **97**(5), 2229–2234.
111. Ezoe S, Matsumura I, Nakata S, *et al*. GATA-2/estrogen receptor chimera regulates cytokine-dependent growth of hematopoietic cells through accumulation of p21(WAF1) and p27(Kip1) proteins. *Blood* 2002, **100**(10), 3512–3520.
112. O'Hagan RC, Ohh M, David G, *et al*. Myc-enhanced expression of Cul1 promotes ubiquitin-dependent proteolysis and cell cycle progression. *Genes Dev* 2000, **14**(17), 2185–2191.
113. Perez-Roger I, Kim SH, Griffiths B, *et al*. Cyclins D1 and D2 mediate myc-induced proliferation via sequestration of p27 (Kip1) and p21 (Cip1. *EMBO J* 1999, **18**(19), 5310–5320.
114. Rothermund K, Rogulski K, Fernandes E, *et al*. C-Myc-independent restoration of multiple phenotypes by two C-Myc target genes with overlapping functions. *Cancer Res* 2005, **65**(6), 2097–2107.
115. Eberhardy SR, Farnham PJ. c-Myc mediates activation of the cad promoter via a post-RNA polymerase II recruitment mechanism. *J Biol Chem* 2001, **276**(51), 48562–48571.
116. Eberhardy SR, Farnham PJ. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *J Biol Chem* 2002, **277**(42), 40156–40162.
117. Kanazawa S, Soucek L, Evan G, *et al*. c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. *Oncogene* 2003, **22**(36), 5707–5711.
118. Felton-Edkins ZA, Kenneth NS, Brown TR, *et al*. Direct regulation of RNA polymerase III transcription by RB, p53 and c-Myc. *Cell Cycle* 2003, **2**(3), 181–184.
119. Gomez-Roman N, Grandori C, Eisenman RN, *et al*. Direct activation of RNA polymerase III transcription by c-Myc. *Nature* 2003, **421**(6920), 290–294.
120. Grummt I. Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev* 2003, **17**(14), 1691–1702.
121. Moss T. At the crossroads of growth control; making ribosomal RNA. *Curr Opin Genet Dev* 2004, **14**(2), 210–217.
122. Arabi A, Wu S, Ridderstrale K, *et al*. c-Myc associates with ribosomal DNA and activates RNA polymerase I transcription. *Nat Cell Biol* 2005, **7**(3), 303–310.
123. Grandori C, Gomez-Roman N, Felton-Edkins ZA, *et al*. c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. *Nat Cell Biol* 2005, **7**(3), 311–318.
124. Grewal SS, Li L, Orian A, *et al*. Myc-dependent regulation of ribosomal RNA synthesis during *Drosophila* development. *Nat Cell Biol* 2005, **7**(3), 295–302.
125. Peukert K, Staller P, Schneider A, *et al*. An alternative pathway for gene regulation by Myc. *EMBO J* 1997, **16**(18), 5672–5686.
126. Roy AL, Carruthers C, Gutjahr T, *et al*. Direct role for Myc in transcription initiation mediated by interactions with TFII-I. *Nature* 1993, **365**(6444), 359–361.
127. Brenner C, Deplus R, Didelot C, *et al*. Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO J* 2005, **24**(2), 336–346.
128. Austen M, Cerni C, Luscher-Firzlaff JM, *et al*. YY1 can inhibit c-Myc function through a mechanism requiring DNA binding of YY1 but neither its transactivation domain nor direct interaction with c-Myc. *Oncogene* 1998, **17**(4), 511–520.
129. Shrivastava A, Saleque S, Kalpana GV, *et al*. Inhibition of transcriptional regulator Yin-Yang-1 by association with c-Myc. *Science* 1993, **262**(5141), 1889–1892.
130. Mao DY, Barsyte-Lovejoy D, Ho CS, *et al*. Promoter-binding and repression of PDGFRB by c-Myc are separable activities. *Nucl Acids Res* 2004, **32**(11), 3462–3468.
131. Marhin WW, Chen S, Facchini LM, *et al*. Myc represses the growth arrest gene gadd45. *Oncogene* 1997, **14**(23), 2825–2834.
132. Barsyte-Lovejoy D, Mao DY, Penn LZ. c-Myc represses the proximal promoters of GADD45a and GADD153 by a post-RNA polymerase II recruitment mechanism. *Oncogene* 2004, **23**(19), 3481–3486.
133. Felsher DW, Bishop JM. Reversible tumourigenesis by MYC in hematopoietic lineages. *Mol Cell* 1999, **4**(2), 199–207.
134. D'Cruz CM, Gunther EJ, Boxer RB, *et al*. c-MYC induces mammary tumourigenesis by means of a preferred pathway involving spontaneous Kras2 mutations. *Nat Med* 2001, **7**(2), 235–239.
135. Shachaf CM, Kopelman AM, Arvanitis C, *et al*. MYC inactivation uncovers pluripotent differentiation and tumour dormancy in hepatocellular cancer. *Nature* 2004, **431**(7012), 1112–1117.
136. Arnold I, Watt FM. c-Myc activation in transgenic mouse epidermis results in mobilization of stem cells and differentiation of their progeny. *Curr Biol* 2001, **11**(8), 558–568.
137. Pelengaris S, Khan M, Evan GI. Suppression of Myc-induced apoptosis in beta cells exposes multiple oncogenic properties of Myc and triggers carcinogenic progression. *Cell* 2002, **109**(3), 321–334.
138. Jonkers J, Berns A. Oncogene addiction: sometimes a temporary slavery. *Cancer Cell* 2004, **6**(6), 535–538.
139. Cotter FE. Unraveling biologic therapy for Bcl-2-expressing malignancies. *Semin Oncol* 2004, **31**(6 Suppl 16), 18–21., discussion 33.
140. Fulda S, Debatin KM. Targeting apoptosis pathways in cancer therapy. *Curr Cancer Drug Targets* 2004, **4**(7), 569–576.
141. Ghobrial IM, Witzig TE, Adjei AA. Targeting apoptosis pathways in cancer therapy. *CA Cancer J Clin* 2005, **55**(3), 178–194.

142. Osford SM, Dallman CL, Johnson PW, *et al*. Current strategies to target the anti-apoptotic Bcl-2 protein in cancer cells. *Curr Med Chem* 2004, **11**(8), 1031–1039.
143. Wickstrom EL, Bacon TA, Gonzalez A, *et al*. Human promyelocytic leukaemia HL-60 cell proliferation and c-myc protein expression are inhibited by an antisense pentadecadeoxynucleotide targeted against c-myc mRNA. *Proc Natl Acad Sci USA* 1988, **85**(4), 1028–1032.
144. Eder PS, DeVine RJ, Dagle JM, *et al*. Substrate specificity and kinetics of degradation of antisense oligonucleotides by a 3′ exonuclease in plasma. *Antisense Res Dev* 1991, **1**(2), 141–151.
145. Crooke ST. Molecular mechanisms of action of antisense drugs. *Biochim Biophys Acta* 1999, **1489**(1), 31–44.
146. Koller E, Gaarde WA, Monia BP. Elucidating cell signalling mechanisms using antisense technology. *Trends Pharmacol Sci* 2000, **21**(4), 142–148.
147. Crooke ST. Molecular mechanisms of antisense drugs: RNase H. *Antisense Nucl Acid Drug Dev* 1998, **8**(2), 133–134.
148. Stahel RA, Zangemeister-Wittke U. Antisense oligonucleotides for cancer therapy – an overview. *Lung Cancer* 2003, **41**(Suppl. 1), S81–S88.
149. Holt JT, Redner RL, Nienhuis AW. An oligomer complementary to c-myc mRNA inhibits proliferation of HL-60 promyelocytic cells and induces differentiation. *Mol Cell Biol* 1988, **8**(2), 963–973.
150. Prochownik EV, Kukowska J, Rodgers C. c-myc Antisense transcripts accelerate differentiation and inhibit G1 progression in murine erythroleukaemia cells. *Mol Cell Biol* 1988, **8**(9), 3683–3695.
151. Carroll JS, Swarbrick A, Musgrove EA, *et al*. Mechanisms of growth arrest by c-myc antisense oligonucleotides in MCF-7 breast cancer cells: implications for the antiproliferative effects of antiestrogens. *Cancer Res* 2002, **62**(11), 3126–3131.
152. Huang Y, Snyder R, Kligshteyn M, *et al*. Prevention of tumour formation in a mouse model of Burkitt's lymphoma by 6 weeks of treatment with anti-c-myc DNA phosphorothioate. *Mol Med* 1995, **1**(6), 647–658.
153. Smith JB, Wickstrom E. Antisense c-myc and immunostimulatory oligonucleotide inhibition of tumourigenesis in a murine B-cell lymphoma transplant model. *J Natl Cancer Inst* 1998, **90**(15), 1146–1154.
154. Smith JB, Wickstrom E. Inhibition of tumourigenesis in a murine B-cell lymphoma transplant model by c-Myc complementary oligonucleotides. *Adv Exp Med Biol* 1998, **451**, 17–22.
155. Citro G, D'Agnano I, Leonetti C, *et al*. c-myc Antisense oligodeoxynucleotides enhance the efficacy of cisplatin in melanoma chemotherapy in vitro and in nude mice. *Cancer Res* 1998, **58**(2), 283–289.
156. Knapp DC, Mata JE, Reddy MT, *et al*. Resistance to chemotherapeutic drugs overcome by c-Myc inhibition in a Lewis lung carcinoma murine model. *Anticancer Drugs* 2003, **14**(1), 39–47.
157. Iversen PL, Arora V, Acker AJ, *et al*. Efficacy of antisense morpholino oligomer targeted to c-myc in prostate cancer xenograft murine model and a Phase I safety study in humans. *Clin Cancer Res* 2003, **9**(7), 2510–2519.
158. Devi GR, Beer TM, Corless CL, *et al*. In vivo bioavailability and pharmacokinetics of a c-MYC antisense phosphorodiamidate morpholino oligomer, AVI-4126, in solid tumours. *Clin Cancer Res* 2005, **11**(10), 3930–3938.
159. Leonetti C, Biroccio A, Benassi B, *et al*. Encapsulation of c-myc antisense oligodeoxynucleotides in lipid particles improves antitumoural efficacy in vivo in a human melanoma line. *Cancer Gene Ther* 2001, **8**(6), 459–468.
160. Zupi G, Scarsella M, Semple SC, *et al*. Antitumour efficacy of bcl-2 and c-myc antisense oligonucleotides in combination with cisplatin in human melanoma xenografts: relevance of the administration sequence. *Clin Cancer Res* 2005, **11**(5), 1990–1998.
161. Stewart DA, Thomas SD, Mayfield CA, *et al*. Psoralen-modified clamp-forming antisense oligonucleotides reduce cellular c-Myc protein expression and B16-F0 proliferation. *Nucl Acids Res* 2001, **29**(19), 4052–4061.
162. Stewart DA, Xu X, Thomas SD, *et al*. Acridine-modified, clamp-forming antisense oligonucleotides synergize with cisplatin to inhibit c-Myc expression and B16-F0 tumour progression. *Nucl Acids Res* 2002, **30**(11), 2565–2574.
163. Pooga M, Land T, Bartfai T, *et al*. PNA oligomers as tools for specific modulation of gene expression. *Biomol Eng* 2001, **17**(6), 183–192.
164. Cutrona G, Carpaneto EM, Ulivi M, *et al*. Effects in live cells of a c-myc anti-gene PNA linked to a nuclear localisation signal. *Nat Biotechnol* 2000, **18**(3), 300–303.
165. Cutrona G, Carpaneto EM, Ponzanelli A, *et al*. Inhibition of the translocated c-myc in Burkitt's lymphoma by a PNA complementary to the E mu enhancer. *Cancer Res* 2003, **63**(19), 6144–6148.
166. Biroccio A, Benassi B, Amodei S, *et al*. c-Myc down-regulation increases susceptibility to cisplatin through reactive oxygen species-mediated apoptosis in M14 human melanoma cells. *Mol Pharmacol* 2001, **60**(1), 174–182.
167. Supino R, Perego P, Gatti L, *et al*. A role for c-myc in DNA damage-induced apoptosis in a human TP53-mutant small-cell lung cancer cell line. *Eur J Cancer* 2001, **37**(17), 2247–2256.
168. Bertrand JR, Pottier M, Vekris A, *et al*. Comparison of antisense oligonucleotides and siRNAs in cell culture and in vivo. *Biochem Biophys Res Commun* 2002, **296**(4), 1000–1004.
169. Wang YH, Liu S, Zhang G, *et al*. Knockdown of c-Myc expression by RNAi inhibits MCF-7 breast tumour cells growth in vitro and in vivo. *Breast Cancer Res* 2005, **7**(2), R220–R228.
170. Song E, Lee SK, Wang J, *et al*. RNA interference targeting Fas protects mice from fulminant hepatitis. *Nat Med* 2003, **9**(3), 347–351.
171. Song E, Zhu P, Lee SK, *et al*. Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. *Nat Biotechnol* 2005, **23**(6), 709–717.
172. Brummelkamp TR, Bernards R, Agami R. Stable suppression of tumourigenicity by virus-mediated RNA interference. *Cancer Cell* 2002, **2**(3), 243–247.
173. McGuffie EM, Catapano CV. Design of a novel triple helix-forming oligodeoxyribonucleotide directed to the major promoter of the c-myc gene. *Nucl Acids Res* 2002, **30**(12), 2701–2709.
174. McGuffie EM, Pacheco D, Carbone GM, *et al*. Antigene and antiproliferative effects of a c-myc-targeting phosphorothioate triple helix-forming oligonucleotide in human leukaemia cells. *Cancer Res* 2000, **60**(14), 3790–3799.
175. Carbone GM, McGuffie E, Napoli S, *et al*. DNA binding and antigene activity of a daunomycin-conjugated triplex-forming oligonucleotide targeting the P2 promoter of the human c-myc gene. *Nucl Acids Res* 2004, **32**(8), 2396–2410.
176. Grand CL, Han H, Munoz RM, *et al*. The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumour growth in vivo. *Mol Cancer Ther* 2002, **1**(8), 565–573.
177. Lemarteleur T, Gomez D, Paterski R, *et al*. Stabilisation of the c-myc gene promoter quadruplex by specific ligands' inhibitors of telomerase. *Biochem Biophys Res Commun* 2004, **323**(3), 802–808.
178. Seenisamy J, Bashyam S, Gokhale V, *et al*. Design and synthesis of an expanded porphyrin that has selectivity for the c-MYC G-quadruplex structure. *J Am Chem Soc* 2005, **127**(9), 2944–2959.
179. Simonsson T, Henriksson M. c-myc Suppression in Burkitt's lymphoma cells. *Biochem Biophys Res Commun* 2002, **290**(1), 11–15.

180. Soucek L, Helmer-Citterich M, Sacco A, *et al*. Design and properties of a Myc derivative that efficiently homodimerises. *Oncogene* 1998, **17**(19), 2463–2472.
181. Soucek L, Jucker R, Panacchia L, *et al*. Omomyc, a potential Myc dominant negative, enhances Myc-induced apoptosis. *Cancer Res* 2002, **62**(12), 3507–3510.
182. Pelengaris S, Littlewood T, Khan M, *et al*. Reversible activation of c-Myc in skin: induction of a complex neoplastic phenotype by a single oncogenic lesion. *Mol Cell* 1999, **3**(5), 565–577.
183. Soucek L, Nasi S, Evan GI. Omomyc expression in skin prevents Myc-induced papillomatosis. *Cell Death Differ* 2004, **11**(9), 1038–1045.
184. Kohlhuber F, Hermeking H, Graessmann A, *et al*. Induction of apoptosis by the c-Myc helix-loop-helix/leucine zipper domain in mouse 3T3-L1 fibroblasts. *J Biol Chem* 1995, **270**(48), 28797–28805.
185. Giorello L, Clerico L, Pescarolo MP, *et al*. Inhibition of cancer cell growth and c-Myc transcriptional activity by a c-Myc helix 1-type peptide fused to an internalization sequence. *Cancer Res* 1998, **58**(16), 3654–3659.
186. Nieddu E, Melchiori A, Pescarolo MP, *et al*. Sequence specific peptidomimetic molecules inhibitors of a protein–protein interaction at the helix 1 level of c-Myc. *FASEB J* 2005, **19**(6), 632–634.
187. Pescarolo MP, Bagnasco L, Malacarne D, *et al*. A retro-inverso peptide homologous to helix 1 of c-Myc is a potent and specific inhibitor of proliferation in different cellular systems. *FASEB J* 2001, **15**(1), 31–33.
188. Berg T, Cohen SB, Desharnais J, *et al*. Small-molecule antagonists of Myc/Max dimerisation inhibit Myc-induced transformation of chicken embryo fibroblasts. *Proc Natl Acad Sci USA* 2002, **99**(6), 3830–3835.
189. Yin X, Giap C, Lazo JS, *et al*. Low molecular weight inhibitors of Myc–Max interaction and function. *Oncogene* 2003, **22**(40), 6151–6159.
190. Guo Y, Cleveland JL, O'Brien TG. Haploinsufficiency for odc modifies mouse skin tumour susceptibility. *Cancer Res* 2005, **65**(4), 1146–1149.
191. Nilsson JA, Keller UB, Baudino TA, *et al*. Targeting ornithine decarboxylase in Myc-induced lymphomagenesis prevents tumour formation. *Cancer Cell* 2005, **7**(5), 433–444.
192. Desbiens KM, Deschesnes RG, Labrie MM, *et al*. c-Myc potentiates the mitochondrial pathway of apoptosis by acting upstream of apoptosis signal-regulating kinase 1 (Ask1) in the p38 signalling cascade. *Biochem J* 2003, **372**(Pt 2), 631–641.
193. Iavarone C, Catania A, Marinissen MJ, *et al*. The platelet-derived growth factor controls c-myc expression through a JNK- and AP-1-dependent signalling pathway. *J Biol Chem* 2003, **278**(50), 50024–50030.
194. Kerkhoff E, Houben R, Loffler S, *et al*. Regulation of c-myc expression by Ras/Raf signalling. *Oncogene* 1998, **16**(2), 211–216.
195. Grumont RJ, Strasser A, Gerondakis S. B cell growth is controlled by phosphatidylinosotol 3-kinase-dependent induction of Rel/NF-κB regulated c-myc transcription. *Mol Cell* 2002, **10**(6), 1283–1294.
196. Watnick RS, Cheng YN, Rangarajan A, *et al*. Ras modulates Myc activity to repress thrombospondin-1 expression and increase tumour angiogenesis. *Cancer Cell* 2003, **3**(3), 219–231.
197. He TC, Sparks AB, Rago C, *et al*. Identification of c-MYC as a target of the APC pathway. *Science* 1998, **281**(5382), 1509–1512.
198. Kiuchi N, Nakajima K, Ichiba M, *et al*. STAT3 is required for the gp 130-mediated full activation of the c-myc gene. *J Exp Med* 1999, **189**(1), 63–73.
199. Oskarsson T, Trumpp A. The Myc trilogy: lord of RNA polymerases. *Nat Cell Biol* 2005, **7**(3), 215–217.

Characterizing cMyc and its interaction with TRRAP

Lilia Kaustov,[1] Sigal Katz,[1] Cynthia S.W. Ho[1#], Romina Ponzielli[1], Shili Duan,[1] Steven B. McMahon[3], Michael D.Cole[4], Linda Z. Penn,[1] and Cheryl H. Arrowsmith[1,2]

[1] Division of Cancer Genomics and Proteomics, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 610 University Ave, Toronto, Ontario M5G 2M9, Canada

[2] Banting and Best Department of Medical Research & Department of Medical Genetics and Microbiology, University of Toronto, 112 College Street, Toronto ON M5G 1L6, Canada

[3] The Wistar Institute, 3601 Spruce Street, Philadelphia, Pennsylvania 19104, USA

[4]Dartmouth College, Molecular and Cellular Biology Program, 7560 Remsen Building, Room 239, Hanover, New Hampshire 03755-3842 USA

 # Present Address: Samuel Lunenfeld Research Institute, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

Corresponding authors who contributed equally to this work:

Cheryl H. Arrowsmith, for structural biology: carrow@uhnres.utoronto.ca

Linda Z. Penn, for molecular biology: lpenn@uhnres.utoronto.ca

Abstract

The N-terminal domain (NTD) of cMyc interacts with specific co-factors to regulate cell cycle, apoptosis and tumorigenesis. TRRAP is one such cofactor that was isolated biochemically by virtue of its interaction with the cMyc NTD. Here we used the Repressed Transactivator (RTA) assay to identify a region of TRRAP (residues 1665-2030) that interacts with the cMyc NTD. Using multiple experimental approaches, we have further defined a region of TRRAP (residues 1690-1850) which is sufficient to bind to a functionally critical region of cMyc (residues 120-160), in a W135 dependent manner both *in vitro* and *in vivo*, establishing the essential components of this critical interaction.

Keywords: TRRAP, cMyc, pulldown assay, circular dichroism, Repressed Transactivator, co-immunoprecipitation

Introduction

Deregulated expression of the c-myc oncogene is a common feature of numerous human cancers of diverse origin and is often associated with aggressive disease. The product of the human c-myc gene, cMyc, is a nuclear phosphoprotein that regulates a wide spectrum of biological activities including cell proliferation, differentiation and programmed cell death. The C-terminal domain of cMyc consists of helix loop helix (HLH) and leucine zipper (LZ) motifs, which are important for cMyc to bind to its obligate partner, Max. The mechanism of cMyc:Max interaction with DNA has been well characterized, in part because the structure of the C-terminal has been solved.[1] By contrast, the biophysical characteristics of the cMyc NTD, alone and in association with its partner proteins has remained elusive, despite the functional importance of this region for all known biological activities of cMyc.

The NTD of cMyc spans amino acids 1 to 262 and contains three regions that are highly conserved amongst all members of the Myc family of proteins including Myc box I (MBI, residues 45- 63), MBII (128-143) and MBIII (188-199).[2] The transactivation domain (TAD), residues 1 to 143, is sufficient to activate gene transcription when linked to a heterologous DNA-binding domain. [3] Analysis of the cMyc TAD by CD has shown that residues 1-143 harbor little or no secondary structural content. [4] Interestingly, structure is evident upon interaction of this region with the basal transcription factor, TATA-binding protein (TBP).[4] Recent analysis of a larger fragment of the cMyc NTD, residues 1-167, which includes the unstructured cMyc TAD, identified a partially helical region and demonstrated features of a molten globule upon specific binding to Myc interactors TBP and MM-1.[5] Thus, evidence suggests the structure and function of the cMyc NTD is highly dependent upon interaction with its partner proteins and operates on an induced fit model of interaction.[6]

One of the most intriguing Myc-binding proteins yet described is the large (~430 KDa) ATM/PI3-kinase-related protein termed, transformation-transactivation domain-associated protein (TRRAP).

Inhibition of TRRAP expression or function blocks Myc-mediated oncogenesis,[7] establishing an essential role for TRRAP in cMyc activity. Indeed, the activation and oncogenic potency of the individual Myc family members was directly proportional to their ability to bind TRRAP.[8] Moreover the interaction between cMyc and TRRAP has been shown *in vivo* to involve the NTD region of cMyc while disruption of a key Tryptophan (W135) residue within MBII was shown to abolish interaction of N-Myc with TRRAP. Taken together, work to date clearly shows that TRRAP holds a critical role in the regulation and function of cMyc and identifying the region of cMyc interaction of this large protein will mark a key advance. [7]

In this study, we utilized a novel RTA technology and identified a large region of TRRAP comprising residues 1665-2030 as important for cMyc interaction. Using CD spectroscopy a folded globular domain of $TRRAP_{1690-1877}$ has been identified and shown to bind $cMyc_{120-160}$ containing the conserved MBII. We found the minimal region of TRRAP comprising residues 1690-1850 exhibited binding to $cMyc_{120-160}$ in a manner that suggests a folding-on-binding mode of interaction between these proteins. This interaction is dependent on W135 in cMyc, a residue that has been shown to be required for Myc-induced transformation activity.[10] Moreover, our immunofluorescence and immunoprecipitation results provide strong evidence of an interaction between cMyc and TRRAP in the nuclear, chromatin-enriched compartment of mammalian cells.

Results and Discussion

Repressed Transactivator Assay defines a cMyc interacting region of TRRAP.

To facilitate the identification of the region within TRRAP that interacts with cMyc, we used an innovative yeast two-hybrid system for bait proteins that have intrinsic transactivation activity. We have previously used this system, the Repressed Transactivator (RTA) assay, to investigate Myc NTD interactors.[11,12] In the RTA (Figure 1(a)), activation of ura3 and lacZ reporter genes by the Gal4-cMyc NTD fusion protein results in the death of the yeast when plated on 5-flouroorotic acid (FOA) media and

elevated expression of beta-galactosidase . The prey proteins in the RTA are fused to the repression domain of the yeast TUP1 protein such that a specific bait:prey interaction is detected as repression of reporter gene expression resulting in the growth of yeast on FOA and inhibition of beta-galactosidase activity.

Using the RTA, several large regions of TRRAP were assayed for their ability to interact with the cMyc NTD bait. The prey vector containing the positive control TBP, the TRRAP fragments I-III (Figure 1(b)), or empty vector (pBDH) were co-transformed with the Gal4-cMyc NTD (Figure 1(c)) into yeast and plated onto media with or without FOA. All yeast were able to grow in the absence of FOA (Figure 1(c), upper panel), but in the presence of FOA (Figure 1(c), lower panel), only TRRAP fragment II (residues 1665-2030) and TBP were able to confer growth-rescue of yeast expressing the Gal4-cMyc NTD. Suppression of lacZ expression to levels similar to those of TBP was observed when TRRAP fragment II was co-transfected into yeast with the cMyc-NTD (Figure 1(d)). Thus, using the RTA we were able to identify a region of TRRAP (1665-2030) that binds the cMyc NTD when expressed in yeast *in vivo*.

We further evaluated the region of cMyc that was essential for $TRRAP_{1665-2030}$ interaction using the RTA. MBII is necessary for Myc transformation activity and Myc interaction with TRRAP.[7] Interestingly a substitution mutation, W135E within MBII, has been shown to be critical for Myc transformation. Using this specific point mutant we evaluated $TRRAP_{1665-2030}$ interaction with the wildtype and W135E cMyc NTD. Growth rescue in the presence of FOA was achieved with expression of $TRRAP_{1665-2030}$ with wildtype cMyc NTD, but significantly diminished in the presence of Gal4-cMycNTD W135E (Figure 1(e)). These data suggest that residue W135 plays a key role in cMyc:TRRAP interaction.

Minimal cMyc binding region

To further define the minimal interaction domains of the cMyc:TRRAP complex, twelve domain-sized fragments, surrounding the cMyc binding region of the $TRRAP_{1690-2026}$ were cloned and tested for expression, solubility and interaction with cMyc (Figure 2(a)). All the fragments were insoluble when

expressed in *E.coli*. To obtain soluble proteins for interaction studies, a refolding protocol (see Materials and Methods) was developed for several TRRAP fragments.

$TRRAP_{1690-2026}$ was expressed at low levels and was proteolytically degraded into a smaller fragment with a molecular weight corresponding to that of $TRRAP_{1690-1899}$ (confirmed by Mass Spectrometry), suggesting the presence of a protease resistant domain. Because of the low expression level of $TRRAP_{1690-1899}$, we designed and cloned a series of smaller $TRRAP_{1690-1899}$ fragments (Figure 2(a)). $TRRAP_{1690-1850}$ and $TRRAP_{1690-1877}$ expressed well, but extensive precipitation precluded refolding of these individual polypeptides (5%-10% refolding yield). However, when both $TRRAP_{1690-1850}$ and $TRRAP_{1690-1877}$ were solubilized with denaturant and refolded in the presence of $cMyc_{120-160}$, these proteins remained soluble suggesting that the folding of these two fragments was facilitated by $cMyc_{120-160}$ (40-50% refolding yield). TRRAP1690-1877 has better solubililty properties and was advanced for further cMyc:TRRAP *in vitro* interactions studies.

While this work was in progress Park et al. described the results of their mapping studies to identify the region of TRRAP that interacts with cMyc and suggested that the minimal region of cMyc interaction may reside in $TRRAP_{1899-2026}$. [12] To directly evaluate whether this region did indeed bind to cMyc, we cloned and expressed this fragment in a similar manner to the other fragments described above. $TRRAP_{1899-2026}$ was the only fragment that could be expressed in soluble form and therefore it was further evaluated for cMyc NTD interaction.

Biophysical analysis of domains

We used CD and NMR spectroscopy to further characterize two cMyc NTD (Figure 2(b)) and TRRAP polypeptides. The CD spectrum of $cMyc_{1-153}$ shows little negative elliplicity at approximately 222 nm and a minimum shifted towards 200 nm which indicates a largely disordered structure with low helical content (Figure 3(a)). A comparison of the CD spectrum of $cMyc_{1-153}$ with that previously published for Myc_{1-143}

indicates similar secondary structural properties. Recent analysis of Myc_{1-167} suggest that residues 140-167 contribute to the folding of the entire cMyc TAD region.[5] This may explain why our $cMyc_{1-153}$ (lacking residues 154-167) had a CD spectrum reflecting significantly less secondary structure than that for Myc_{1-167}. The CD spectra of $cMyc_{120-160}$ on the other hand displays negative minima near 208 nm and 222 nm (Figure 3(a)), consistent with some α helical content and is similar to CD spectra of Myc_{1-167} and Myc_{92-167}.[5] This region contains MBII residues 128-143 which are conserved in all Myc family proteins and known to be essential for the cMyc:TRRAP interaction.

The [^{1}H-^{15}N]HSQC spectrum of $cMyc_{120-160}$ (Figure 3(c)) a soluble, partially folded protein with the approximate number of expected cross peaks. The non-uniform peak intensity and lack of chemical shift dispersion are both consistent with a dynamically disordered and/or partially helical polypeptide as suggested by the CD spectra. The HSQC of $cMyc_{1-153}$ (not shown) reflects a primarily disordered protein with extensive spectral overlap. Therefore, $cMyc_{120-160}$, was advanced for further cMyc:TRRAP interaction studies.

The Ul-^{15}N-$TRRAP_{1899-2026}$ HSQC spectrum reflects a soluble, non-aggregated, but disordered protein with very little chemical shift dispersion, and uniformly narrow peak width (Fig. 3(d)). Addition of $cMyc_{1-153}$ had no effect on the spectral features of Ul-^{15}N-$TRRAP_{1899-2026}$ (data not shown). We used GST pulldown assay to (see below) to evaluate the interaction of $TRRAP_{1899-2026}$ with $cMyc_{120-160}$. No interaction was observed between these proteins. Therefore, we conclude that there is very little or no interaction between $TRRAP_{1899-2026}$ and both cMyc fragments under these conditions, as a complex could not be identified by either NMR, or by GST pull down assays at millimolar protein concentrations.

Using CD spectroscopy we have identified a region of TRRAP that appears to have a folded, globular domain ($TRRAP_{1690-1877}$) and is likely to form a complex with the MBII. While the spectrum of $cMyc_{120-160}$ alone was characteristic of a mix of helical and random coil conformation, the CD spectrum of $TRRAP_{1690-}$

$_{1877}$ fragment alone had a strong minimum near 217 nm, characteristic of a β-sheet structure. The spectrum of a stoichiometric mixture of the two proteins is clearly distinct from those of $TRRAP_{1690-1877}$ and $cMyc_{120-160}$ alone, and from the sum of the two individual CD spectra with an obvious minimum at 222 nm (Fig. 3(b)) indicative of the conversion of a random coil to an α-helical conformation, and consistent with an interaction between these domains.

In addition, we focused on a particular residue within MBII, W135, that has been shown to be required for Myc-induced transformation. Interestingly when the $cMyc_{120-160}$ mutant , W135E, was mixed in stoichiometric amounts with $TRRAP_{1690-1877}$ the CD spectrum of the mixture showed no difference from the sum of the individual spectra suggesting no interaction to place (Figure 3b).

In vitro binding confirms cMyc:TRRAP minimal domains

A more direct assay of protein:protein interactions using the *in vitro* GST pulldown technique was performed to evaluate the interaction of between cMyc and TRRAP. Immobilized GST-$cMyc_{120-160}$ was incubated with purified His-tagged $TRRAP_{1690-1877}$, $TRRAP_{1690-1850}$ or $TRRAP_{1899-2021}$. The GST-$cMyc_{120-160}$W135E was incubated with His tagged $TRRAP_{1690-1877}$. The results clearly showed an interaction of $cMyc_{120-160}$ with both $TRRAP_{1690-1877}$ and $TRRAP_{1690-1850}$, but not with $TRRAP_{1899-2021}$ (Figure 4). No interaction was observed between $cMyc_{120-160}$W135E and $TRRAP_{1690-1877}$ (Figure 4(d)). Together, these data indicate that $cMyc_{120-160}$ can bind to $TRRAP_{1690-1850}$ and $TRRAP_{1690-1877}$ and a nonconservative mutation within MBII, W135E, was able to disrupt the binding between these proteins. This is consistent with our RTA (Fig. 1(e)) and the CD spectra which could be explained by either a destabilizing affect on the structural integrity of MBII, or direct participation of W135 in the cMyc:TRRAP interaction. Thus, consistent with previous reports, we conclude that MBII sequence is required for efficient interaction of the cMyc NTD and TRRAP.

$TRRAP_{1690-1850}$ interacts with cMyc in vivo.

We further investigated whether the interaction of $TRRAP_{1690-1850}$ and cMyc was detectable in mammalian cells *in vivo*. To this end, 293TV cells were transiently transfected with pCMV10 expression plasmid carrying a fusion protein containing a FLAG-tagged nuclear localization signal (NLS) linked in-frame with $TRRAP_{1690-1850}$. To investigate the sub-cellular localization of $TRRAP_{1690-1850}$, we conducted immunofluorescence studies and biochemical analyses. Cells were fixed and immuno-stained with polyclonal anti-Myc and monoclonal anti-FLAG antibodies to detect Myc and $TRRAP_{1690-1850}$ respectively. As shown in Figure 5(a-d), $TRRAP_{1690-1850}$ and cMyc primarily co-localize in the nucleus of 293TV cells, in a diffuse dot pattern (white arrows in Figure 5(c)). In addition, a less prevalent staining of $TRRAP_{1690-1850}$ is shown in the cytoplasmic compartment, most likely due to over-expression of the protein. For biochemical analysis, cell lysates were harvested and sub-cellular fractions prepared. Equivalent volumes of cytoplasmic, nuclear and chromatin-enriched fractions were separated on a 12% SDS-PAGE gel (Figure 5(e)). In addition to anti-FLAG, immunoblots were probed with anti-tubulin and anti-acetylated histone H3 to monitor efficiency of cellular fractionation. Interestingly, Flag-NLS-$TRRAP_{1690-1850}$ was highly enriched in the chromatin fractions, yet also evident in the cytoplasmic fraction, in accordance with the immunofluorescence results. To determine if $TRRAP_{1690-1850}$ interacts with full length cMyc *in vivo* after transient co-expression in 293TV cells we conduced co-immunoprecipitation experiments. Cell lysates were immunoprecipitated with anti-Myc antibodies (N-262) followed by immunoblotting with anti-FLAG antibodies to identify FLAG-$TRRAP_{1690-1850}$ protein (Figure 5(f)). As shown in Figure 5(f) lane 4, FLAG-$TRRAP_{1690-1850}$ was specifically co-immunoprecipitated by cMyc and not by rabbit IgG (lane 3).

Our immunofluorescence and immunoprecipitation results provide strong evidence of an interaction between cMyc and TRRAP in the nuclear, chromatin-enriched compartment of mammalian cells. Indeed, ectopically expressed TRRAP colocalizes with cMyc in 293TV cells. As well $TRRAP_{1690-1850}$ resides in the

chromatin-enriched fraction, along with cMyc, and is co-immunoprecipitated with ectopic cMyc from mammalian cells.

Summary

In this study, we provide evidence of a physical interaction between cMyc and $TRRAP_{1690-1877}$ from analyses conducted both *in vitro* (CD, GST-pulldown) and *in vivo* (co-IF and co-IP in mammalian cells). The present results thus suggest that $cMyc_{120-160}$ contributes significantly both to folding and binding properties, and may be required for full biophysical and biological functionality of the cMyc TAD region. Importantly, our finding of increased secondary structure in the complex compared to that of the individual polypeptides and the improved behavior of $TRRAP_{1690-1877}$ in the presence of $cMyc_{120-160}$ suggest an induced-fit model of cMyc NTD protein:protein interaction.

Our work extends a previous report using GST pull down assays to show that cMyc and $TRRAP_{1591-2026}$ can interact, but are inconsistent with previous co-immunoprecipitation data showing cMyc and $TRRAP_{1899-2401}$ can interact. It may be that another cMyc-binding region resides in this latter portion of TRRAP, however our results are not supportive of this possibility. We show that cMyc does not interact with $TRRAP_{1899-2026}$ *in vitro*, and a region of TRRAP corresponding to residues 2000-2365 (Figure 1) did not interact with the cMyc NTD in the yeast RTA assay *in vivo*. Our preferred model is that the weak interaction observed between cMyc and $TRRAP_{1899-2401}$ in mammalian cells may be a consequence of indirect binding of cMyc to another transcriptional activator that binds TRRAP in this region. Indeed, the large TRRAP protein interacts with several transcriptional activators through distinct regions of interaction and has been suggested to bridge functional binding interactions between different activators.[13] It will be interesting to further dissect which activator(s) binds to $TRRAP_{1899-2401,}$ as previous results have shown ectopic expression of this region can prolong the doubling time of IRM-5 neuroblastoma cells. For example,

p53 has been shown to bind $TRRAP_{1992-2370}$[14] and may be one of many candidate activators that bind $TRRAP_{1899-2401}$ to repress cellular proliferation.

The specific interaction of $cMyc_{120-160}$ and $TRRAP_{1690-1877}$ provides a novel protein:protein interaction that is an attractive target for therapeutic intervention in cMyc activated cancers. These results establish the basis for subsequent biochemical and structural analyses of the cMyc:TRRAP complex leading to a more comprehensive view of the cMyc regulatory network and will provide new insights into understanding how cMyc function is regulated by interaction with the transcriptional coactivator TRRAP.

Materials and Methods

Protein expression and purification

All proteins were expressed in *E. coli* BL21-CodonPlus (DE3)-RIL cells (Stratagene) and purified using metal affinity chromatography under native conditions for the $TRRAP_{1899-2026}$ and in the presence of guanidinium chloride (GnCl) for cMyc and other TRRAP fragments. The latter underwent subsequent refolding by rapid dilution and dialysis against 25mM TRIS pH 8, 500mM NaCl, 2 mM $MgCl_2$, 10% Glycerol, 0.55M GnCl followed by dialysis against the same buffer without glycerol and GnCl. $cMyc_{120-160}$ and $cMyc_{1-153}$ were refolded by 2-step dialysis in 25mM NaH_2PO_4, 150mM NaCl, pH 6.5, 3M GnCl, followed by dialysis against the same buffer without GnCl. All buffers contained 2mM benzamidine, 0.5mM PMSF and 0.5mM TCEP.

GST pull down assay

Purified $TRRAP_{1899-2026}$, $TRRAP_{1690-1850}$ and $TRRAP_{1690-1877}$ HIS-tagged proteins were incubated with GST-fusion $cMyc_{120-160}$ and $cMyc_{120-160}$(W135E) bound to the GST beads at 4^oC for 1hr. After extensive washing with assay buffer, bound proteins were eluted with 20mM reduced glutathione and detected by SDS-PAGE gel. Control assays with GST protein were also completed to validate results.

CD and NMR spectroscopy

CD scans were performed on AVIV 62DS spectrometer. Protein concentrations used for CD were 0.1-0.15mg/ml. An equimolar solution of $cMyc_{120-160}$ and $TRRAP_{1690-1877}$ was used for the mixed spectra. The mean residue elliplicity (θ) is expressed in units of 10^3 deg. cm^2 $dmol^{-1}$.

All NMR spectra were recorded at 25°C on Varian INOVA 500 MHz spectrometer equipped with room temperature triple-resonance probe. The final NMR samples contained 90%H_2O/10%D_2O with a protein concentration 0.2-0.4 mM.

Immunoprecipitation and Immunoblot

Co-immunoprecipitation was preformed using 1mg of whole cell lysates prepared using lysis buffer A (15mM Tris-HCL pH 7.5, 500mM Nacl, 0.35% NP-40, 5mM EGTA, 5mM EDTA) and incubated over-night with 10µl of anti-Myc (rabbit sera) or IgG (Santa Cruz) and G-sepharose beads (Santa Cruz) .. After 5 washes in Buffer B (25mM Tris-HCL pH 7.5, 250mM Nacl, 0.5% NP-40, 1mM EDTA, 10% glycerol, 1mM DTT) immunoprecipitates were resolved by 12% SDS-PAGE and subjected to immunoblot analysis using anti-Myc (9E10) and anti-FLAG monoclonal antibodies (M2; Sigma).

Chromatin association assay

293TV cell transiently expressing pCMV10-FLAG- $TRRAP_{1690-1850}$ were processed as described by Wysocka et al.[15]

Immunofluorescence

Transiently transfected 293TV cells were fixed in 3% paraformaldehyde, permeabilized in 0.1% Tween 20/PBS, and incubated with anti-Myc (1:500, N-262) or anti-FLAG (1:50, M2) antibodies. Cells were washed, incubated with Alexa-488 anti-mouse or Alexa-568 anti-rabbit (Molecular Probes, Inc.) secondary antibodies and processed for confocal microscopy.

Acknowledgements

This research was supported by the National Cancer Institute of Canada (NCIC) with funds from the Canadian Cancer Society (C.A.) and the U.S. Department of Defense Breast Cancer Research Program grant BC032138 (L.P.). L.K. is a recipient of CIHR, LLS and ISF fellowships. S.K. and R.P. are the recipients of NCIC and the Knudson Postdoctoral Fellowship Award, respectively.

References

[1] Nair, S.K. and Burley, S.K. (2006) Structural aspects of interactions within the Myc/Max/Mad network,Curr Top Microbiol Immunol 302, 123-43.

[2] Oster, S.K., Ho, C.S., Soucie, E.L. and Penn, L.Z. (2002) The myc oncogene: MarvelouslY Complex,Adv Cancer Res 84, 81-154.

[3] Kato, G.J., Barrett, J., Villa-Garcia, M. and Dang, C.V. (1990) An amino-terminal c-myc domain required for neoplastic transformation activates transcription,Mol Cell Biol 10, 5914-20.

[4] McEwan, I.J., Dahlman-Wright, K., Ford, J. and Wright, A.P. (1996) Functional interaction of the c-Myc transactivation domain with the TATA binding protein: evidence for an induced fit model of transactivation domain folding,Biochemistry 35, 9584-93.

[5] Fladvad, M., Zhou, K., Moshref, A., Pursglove, S., Safsten, P. and Sunnerhagen, M. (2005) N and C-terminal sub-regions in the c-Myc transactivation region and their joint role in creating versatility in folding and binding,J Mol Biol 346, 175-89.

[6] Hermann, S., Berndt, K.D. and Wright, A.P. (2001) How transcriptional activators bind target proteins,J Biol Chem 276, 40127-32.

[7] McMahon, S.B., Van Buskirk, H.A., Dugan, K.A., Copeland, T.D. and Cole, M.D. (1998) The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins,Cell 94, 363-74.

[8] Nikiforov, M.A., Chandriani, S., Park, J., Kotenko, I., Matheos, D., Johnsson, A., McMahon, S.B. and Cole, M.D. (2002) TRRAP-dependent and TRRAP-independent transcriptional activation by Myc family oncoproteins,Mol Cell Biol 22, 5054-63.

[9] Wood, M.A., McMahon, S.B. and Cole, M.D. (2000) An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc,Mol Cell 5, 321-30.

[10] Oster, S.K., Mao, D.Y., Kennedy, J. and Penn, L.Z. (2003) Functional analysis of the N-terminal domain of the Myc oncoprotein,Oncogene 22, 1998-2010.

[11] Hirst, M., Ho, C., Sabourin, L., Rudnicki, M., Penn, L. and Sadowski, I. (2001) A two-hybrid system for transactivator bait proteins,Proc Natl Acad Sci U S A 98, 8726-31.

[12] Park, J., Kunjibettu, S., McMahon, S.B. and Cole, M.D. (2001) The ATM-related domain of TRRAP is required for histone acetyltransferase recruitment and Myc-dependent oncogenesis,Genes Dev 15, 1619-24.

[13] Cheng, A.S. et al. (2006) Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters,Mol Cell 21, 393-404.

[14] Ard, P.G., Chatterjee, C., Kunjibettu, S., Adside, L.R., Gralinski, L.E. and McMahon, S.B. (2002) Transcriptional regulation of the mdm2 oncogene by p53 requires TRRAP acetyltransferase complexes,Mol Cell Biol 22, 5650-61.

[15] Wysocka, J., Reilly, P.T. and Herr, W. (2001) Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells,Mol Cell Biol 21, 3820-9.

Figure captions

Figure 1. $TRRAP_{1690-2030}$ interacts with Gal4-Myc NTD. (a) Interaction between the bait and prey in the RTA results in repression of reporter genes ura3 and lacZ as detected by yeast cell growth on FOA and inhibition of beta-galactosidase activity, respectively. (b) Schematic representation of TRRAP showing relative position of the fragments analyzed in the RTA. (c) Gal4-Myc NTD bait was co-transformed into yeast cells with TBP (positive control) or TRRAP fragments or the repression domain alone (pBDH). Resulting colonies were plated onto +/-FOA media. (d) Yeast co-transfected as in (c) were evaluated for *lacZ* expression. (e) Yeast co-transfected with Gal4-Myc NTD wildtype (wt) or Gal4-Myc-NTDW135E with TRRAP fragment II were evaluated for growth in the presence of FOA.

Figure 2. $TRRAP_{1690-2026}$ central region expression constructs with indicated expression levels: (-) no expression , (+) low level (1mg/1L LB), and (+++) good level of expression (8-12mg/1L LB) (b) Schematic representation of human cMyc highlighting the relative location of the TAD as well as the conserved domains within the NTD. The NTD expression constructs are also indicated.

Figure 3. Secondary structure characterization of cMyc and TRRAP proteins: (a) CD spectra of $cMyc_{1-153}$ and $cMyc_{120-160}$. (b) Interaction of $cMyc_{120-160}$ with TRRAP causes a conformational change. The CD spectrum was measured for $TRRAP_{1690-1877}$ and $cMyc_{120-160}$ alone and for a mixture of two proteins, $TRRAP_{1690-1877}/cMyc_{120-160}$, as well as for $TRRAP_{1690-1877}/cMyc_{120-160}$W135E. [^{15}N-^{1}H]HSQC spectrum of: (c) $cMyc_{120-160}$ and (d) $TRRAP_{1899-2026}$.

Figure 4. $TRRAP_{1690-1877}$ and $TRRAP_{1690-1850}$ mediate $cMyc_{120-160}$ binding. Immobilized GST- $cMyc_{120-160}$ co-incubated with His-tagged: (a) $TRRAP_{1899-2026}$; (b) $TRRAP_{1690-1877}$; and (c) $TRRAP_{1690-1877}$ polypeptides (L). (d) Immobilized GST-$Myc_{120-160}$W135E mutant co-incubated with $TRRAP_{1690-1877}$ (L). After washing with PBSx1 buffer and elution with reduced glutatione buffer (E), interacting proteins were subjected to SDS PAGE analysis. (*) indicates degradation products.

Figure 5. $TRRAP_{1690-1850}$ interacts with c-Myc in the nucleus of mammalian cells. 293TV cells (a-d) were transiently transfected with FLAG-NLS-$TRRAP_{1690-1850}$ and c-Myc, Cells were stained with anti-FLAG antibody and anti-Myc antibodies and processed for confocal imaging. (a) Green immunofluorescence staining shows nuclear and some cytoplasmic localization of $TRRAP_{1690-1850}$. (b) Nuclear localization of c-Myc was shown by red fluorescence. (c). Merged fluorescence images indicate nuclear co-localization of $TRRAP_{1690-1850}$ and c-Myc (white arrows). (d) Dapi staining (e) 293TV cells transfected with FLAG-NLS-$TRRAP_{1690-1850}$ plasmid were harvested and sub-cellular fractions prepared. Equivalent volume subfractions were separated by SDS-PAGE. The blot was probed with anti-FLAG then re-probed with fractionation controls, anti-Histone H3 and anti-alpha tubulin antibodies. (f) 293TV cells were transfected with FLAG-NLS-$TRRAP_{1690-1850}$ and c-Myc. Total cell extract was subjected to immunoprecipitation with anti-c-Myc antibodies or rabbit IgG. Co-immunoprecipitated $TRRAP_{1690-1850}$ was visualized after SDS-PAGE and immunoblotting with anti-FLAG antibodies (bottom). The upper part of the blot was probed with anti-Myc antibodies (top); Input control (10%) is shown in the second lane while a cell extract of mock transfected cells is shown in the first lane.

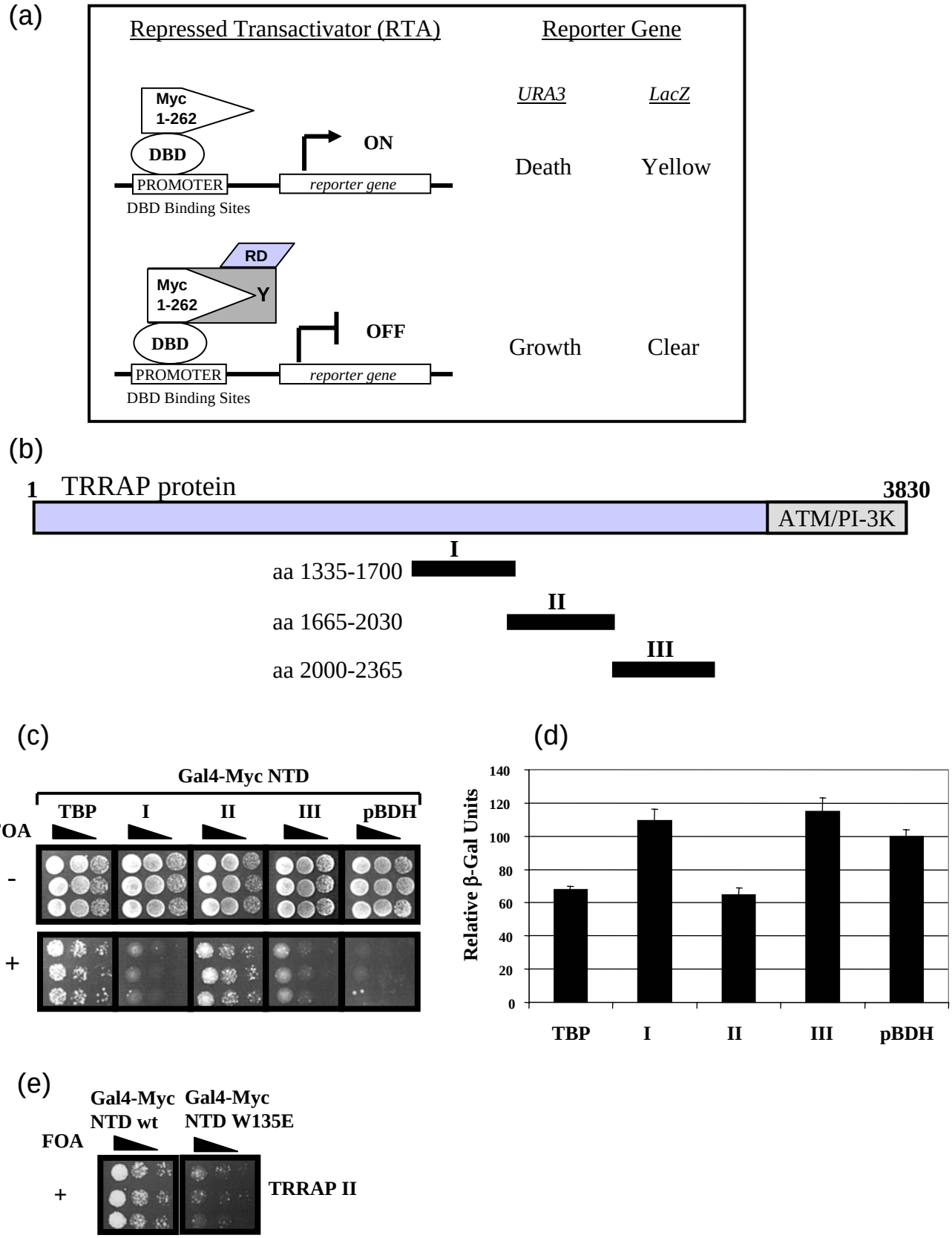

(a)
Repressed Transactivator (RTA)
Reporter Gene
URA3
LacZ
Myc
1-262
DBD
ON
PROMOTER
reporter gene
DBD Binding Sites
Death
Yellow
RD
Myc
1-262
Y
DBD
OFF
PROMOTER
reporter gene
DBD Binding Sites
Growth
Clear
(b)
1
TRRAP protein
3830
ATM/PI-3K
I
aa 1335-1700
II
aa 1665-2030
III
aa 2000-2365
(c)
Gal4-Myc NTD
TBP
I
II
III
pBDH
FOA
-
+
(d)
Relative β-Gal Units
140
120
100
80
60
40
20
0
TBP
I
II
III
pBDH
(e)
Gal4-Myc
NTD wt
Gal4-Myc
NTD W135E
FOA
+
TRRAP II

Figure 1.

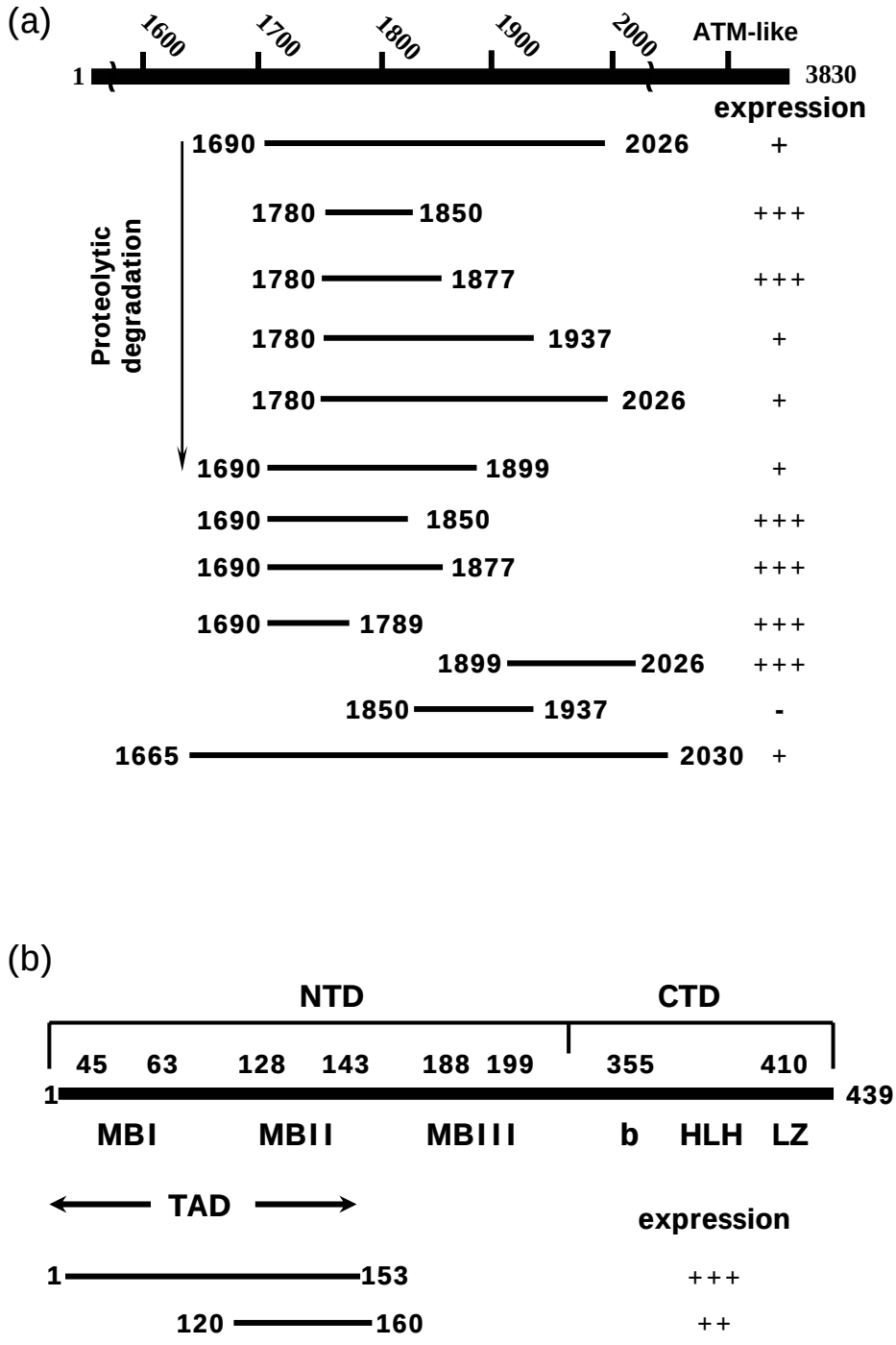

(a)
1600
1700
1800
1900
2000
ATM-like
1
3830
expression
1690 2026 +
Proteolytic degradation
1780 1850 +++
1780 1877 +++
1780 1937 +
1780 2026 +
1690 1899 +
1690 1850 +++
1690 1877 +++
1690 1789 +++
1899 2026 +++
1850 1937 -
1665 2030 +
(b)
NTD
CTD
45 63 128 143 188 199 355 410
1
439
MBI MBII MBIII b HLH LZ
TAD
expression
1 153 +++
120 160 ++

Figure 2.

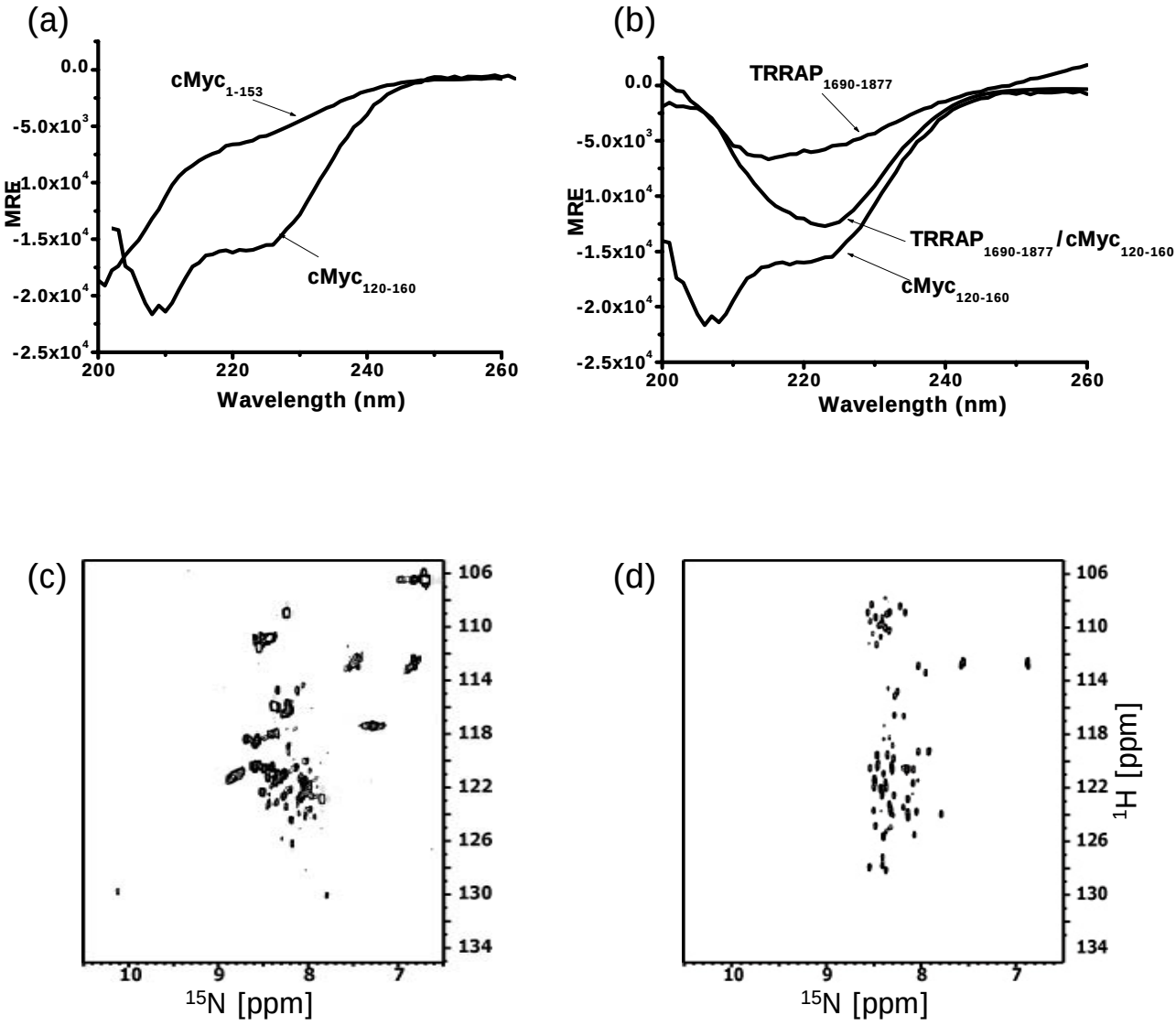

(a)
cMyc1-153
cMyc120-160
MRE
Wavelength (nm)
(b)
TRRAP1690-1877
TRRAP1690-1877/cMyc120-160
cMyc120-160
MRE
Wavelength (nm)
(c)
15N [ppm]
(d)
15N [ppm]
1H [ppm]

Figure 3.

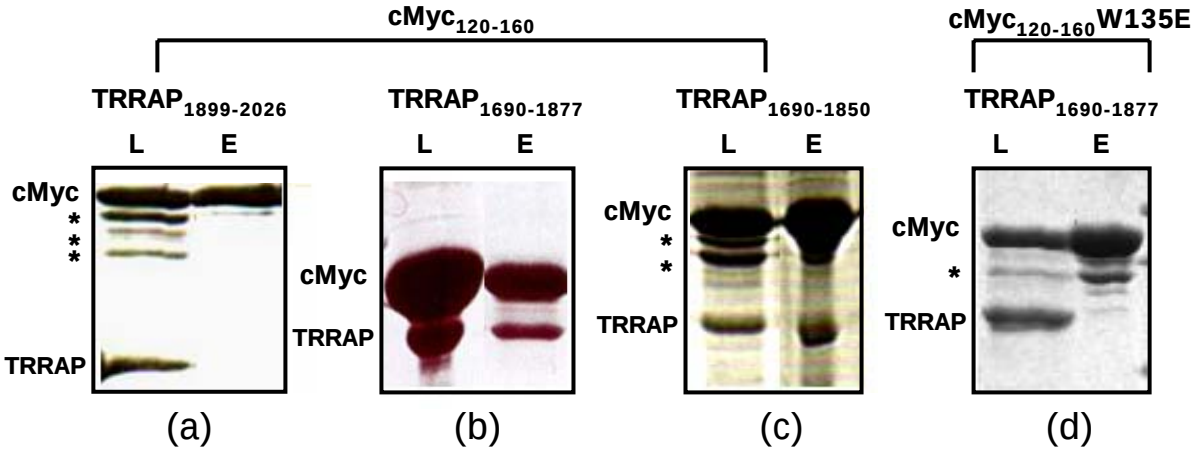
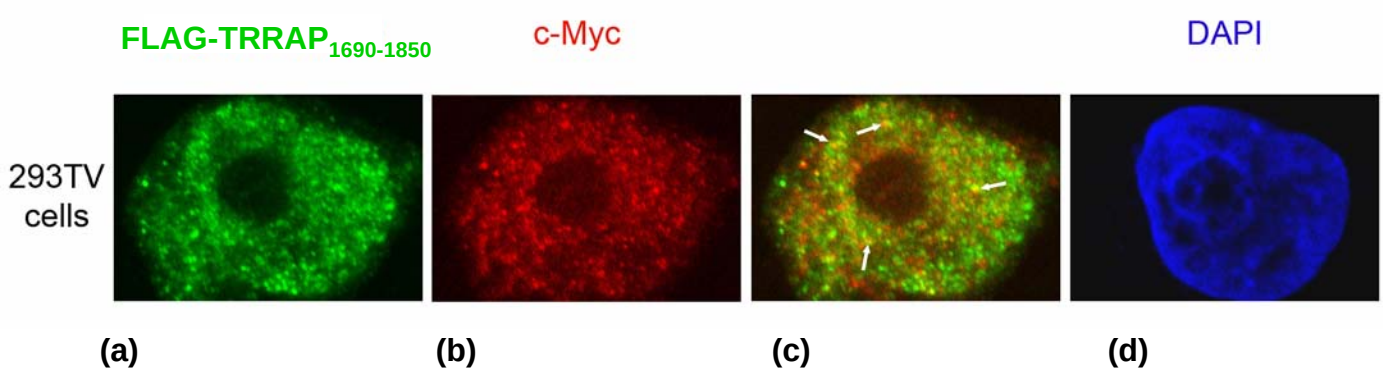
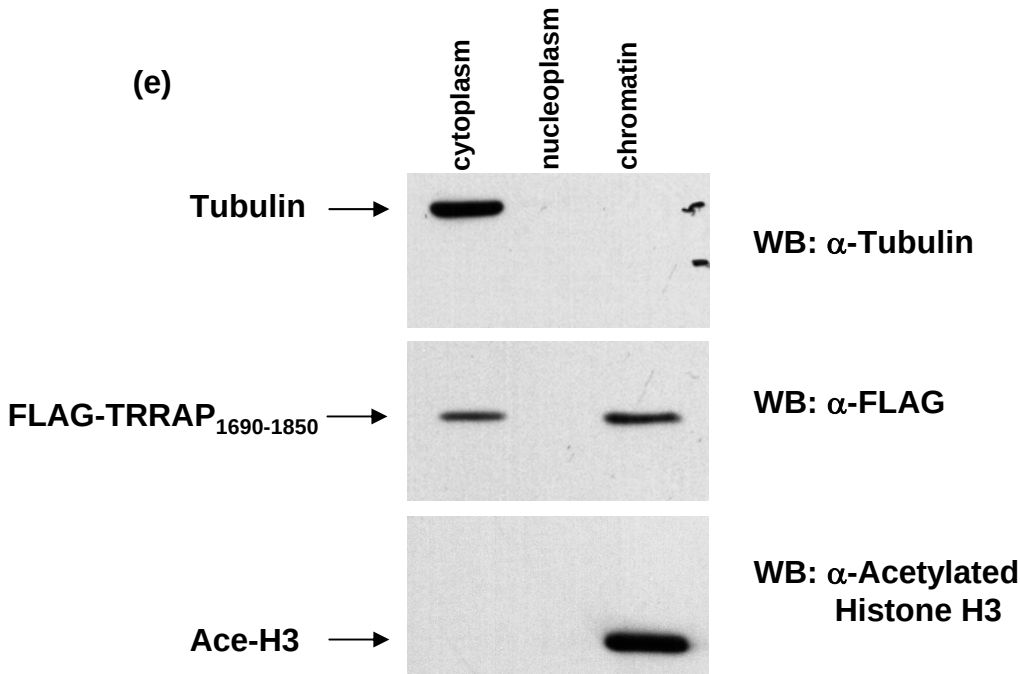

cMyc120-160
cMyc120-160W135E
TRRAP1899-2026
TRRAP1690-1877
TRRAP1690-1850
TRRAP1690-1877
L
E
cMyc
TRRAP
(a)
(b)
(c)
(d)

Figure 4.


FLAG-TRRAP1690-1850
c-Myc
DAPI
293TV cells
(a)
(b)
(c)
(d)


(e)
cytoplasm
nucleoplasm
chromatin
Tubulin
WB: α-Tubulin
FLAG-TRRAP1690-1850
WB: α-FLAG
Ace-H3
WB: α-Acetylated Histone H3

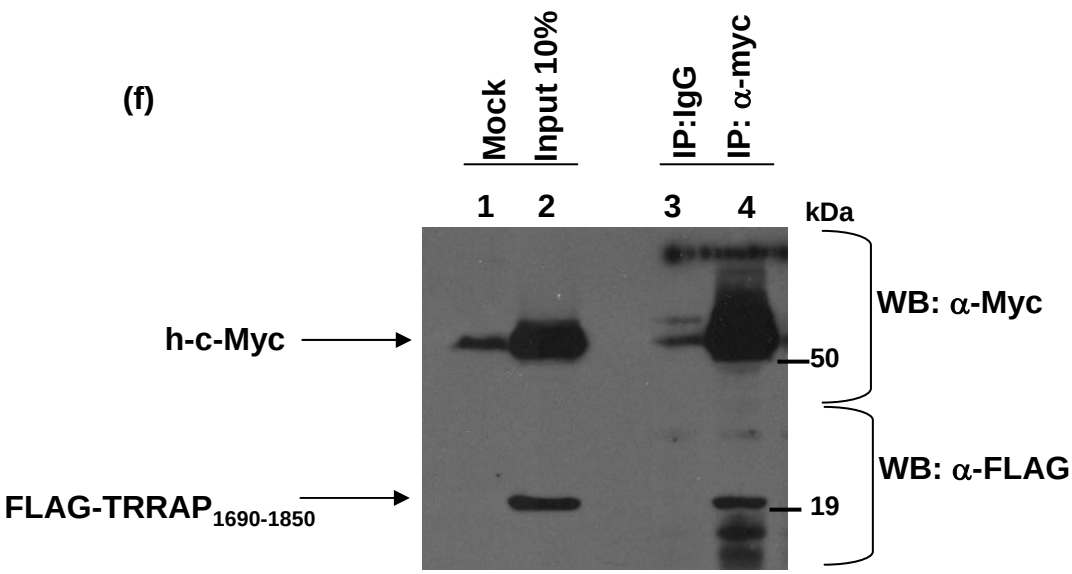

(f)
Mock
Input 10%
IP:IgG
IP: α-myc
1
2
3
4
kDa
h-c-Myc
WB: α-Myc
50
FLAG-TRRAP1690-1850
WB: α-FLAG
19

Figure 5

Optimization of experimental design parameters for high-throughput chromatin immunoprecipitation studies

Romina Ponzielli[1,5], Paul C. Boutros[1,2,3,5], Sigal Katz[1], Angelina Stojanova[1,3], Adam P. Hanley[1,3], Fereshteh Khosravi[1], Igor Jurisica[2,3,4], Linda Z. Penn[1,3,*]

[1]Division of Cancer Genomics and Proteomics, [2]Division of Signaling Biology, Ontario Cancer Institute, University Health Network, Toronto, ON, M5G 2M9, Canada

[3]Department of Medical Biophysics, [4]Department of Computer Science, University of Toronto, ON, M5S 1A8, Canada

[5]Both authors contributed equally to the work

[*]Corresponding author

lpenn@uhnres.utoronto.ca (L.Z. Penn)

Dr. Linda Z. Penn
Ontario Cancer Institute
Room 9-625
610 University Avenue
Toronto, Ontario
M5G 2M9
Phone number: (416) 946-2276
Fax Number: (416) 946-2840

e-mail addresses: rponziel@uhnres.utoronto.ca, Paul.Boutros@utoronto.ca, katz_sigal@yahoo.com, astojano@uhnres.utoronto.ca, adam.hanley@utoronto.ca, khosravi@uhnres.utoronto.ca, juris@ai.utoronto.ca

ABSTRACT

High-throughput, microarray-based chromatin immunoprecipitation (ChIP-chip) technology, allows *in vivo* elucidation of transcriptional networks. However, this complex assay is not readily accessible, in part because its many parameters have not been systematically evaluated and optimized. We fill this gap by addressing experimental-design parameters including antibody purity, dye-bias, array-batch, inter-day hybridization, amplification method, and choice of hybridization control. The combined performance of these optimized parameters shows high validation rates in a ChIP-chip analysis of Myc genomic binding in HL60 cells. Increased sensitivity and decreased noise in ChIP-chip assays will enable wider use of this methodology to accurately and affordably elucidate transcriptional networks.

BACKGROUND

The combination of Chromatin immunoprecipitation (ChIP) with high-throughput DNA microarray technology (chip) provides a powerful method (ChIP-chip) for mapping protein-DNA interactions *in vivo*. The ChIP-chip procedure was first used to study yeast transcription factors [1-4], and has recently been exploited for similar studies in mammalian cells [5-14]. In the ChIP-chip procedure genomic DNA is precipitated in ChIP reactions using specific and control antibodies, then labeled and hybridized to genomic microarrays. Several different array platforms have been used, including proximal promoter region arrays [11], CpG island arrays [12, 15] and whole-genome tiling arrays [16, 17]. ChIP-chip is an extremely versatile technique, and is capable of identifying target genes whose regulatory regions are bound by transcription factors through protein-DNA and/or protein-protein interactions. Moreover, ChIP-chip can be applied to any cell or tissue, making it possible to decode gene regulatory networks *in vivo*. Certainly, the ability to profile *bona fide* target genes has been revolutionized by ChIP-chip. Unlike mRNA expression analysis, the genetic program specifically directed by the transcription factor can be distinguished from subsequent downstream regulatory events. [18-22].

Despite ChIP-chip methodology having become the gold standard for discovering the genomic binding sites of transcriptional regulators, there is wide variability in experimental design, which complicates and delays widespread application. ChIP-chip experiments require careful optimization of a number of parameters. For exemple, the number of cells or amount of tissue used as starting material varies widely from one

study to another [23, 24]. Next, the protein-protein and/or protein-DNA cross-linking, chromatin sonication, as well as antibody sensitivity and purity characteristics can also vary significantly. In most studies, the enriched DNA recovered after the ChIP procedure is amplified, using a variety of methods such as ligation-mediated [24, 25], random priming [26] and T7 priming [27] PCR. Finally, when the amplified and labeled DNA is hybridized to a microarray, a control sample must be selected, and the effects of array batch and dye-swap status considered.

Extensive evaluation of experimental design parameters for mRNA expression arrays have been performed by a number of groups over the past decade. As a result, the key factors are well understood and the assay has been optimized [28]. It is possible, for example, to estimate the number of biological replicates required to sufficiently power a specific hypothesis-testing question [29]. Despite clear evidence that parameter optimization can profoundly impact array analysis, ChIP-chip design parameters have not yet been thoroughly and systematically investigated.

Here, we fill that gap by providing a comprehensive evaluation of experimental design parameters for ChIP-chip studies. Through a series of validation studies, we address both the parameters previously investigated for mRNA expression studies as well as those specific to ChIP-chip experiments. We exploit a well-characterized system: the genomic binding of the Myc oncoprotein in HL60 cells, a human myelogenous leukemia cell line, combined with CpG island arrays [12, 30]. Many parameters for successful ChIP-chip studies were analyzed, including the impact of antibody purity, array batch variability, dye-bias, inter-day hybridization-variability, amplification procedure, and hybridization control. In addition, we evaluated the combined effect of the optimized

parameters by conducting a Myc ChIP-chip study, which showed a high rate of validation by real time PCR. The raw data from this study, encompassing nearly 100 arrays, will be deposited in the GEO repository at NCBI. Our careful description of ChIP-chip experimental design is a key step towards enabling widespread exploitation of this important technology for the rapid elucidation of transcriptional regulatory networks.

RESULTS

Our analysis of ChIP-chip design parameters was conducted in the context of the genomic-binding of the well characterized Myc oncogene in an acute myelogenous leukemia cell line [12], HL60 [31]. All experiments utilized 12k CpG island microarrays [30].

Antibody selection, array batch, reciprocal labeling and inter-day processing

The first series of experiments focused on assessing four key parameters of the ChIP-chip assay. Antibody selection was considered first because the immunoprecipitation step provides the starting material for the array analyses, and as such is a major determinant of the success of ChIP-chip studies. Second, we addressed the effect of array batch. This can be a major confounding variable because DNA spotted to a microarray slide may degrade or otherwise lose sensitivity, and even arrays printed in different print-runs from a single spotter can show variability. Third, we evaluated the bias introduced during the process of sample labeling. This bias refers to intensity differences between identical samples labeled with different dyes and can arise in several ways, including differential incorporation efficiencies for the two dyes, fluorophore degradation, as well as photo-bleaching rates or scanner-induced bias. Fourth, we determined whether day-to-day variations in atmospheric ozone, technical handling, scanner calibration, or other, as yet unidentified, confounders affect the hybridization step.

To evaluate the effect of these ChIP-chip parameters, we conducted a saturated,

fully blocked experiment (Fig. 1). First, six independent biological replicates of cross-linked HL60 cells were subjected to ChIP with either commercial or home-made polyclonal antibodies raised against the N-terminal 262 residues of the Myc oncoprotein. For every replicate ChIP reaction we probed two batches of arrays separated in print-age by about 6 months to evaluate array batch effect. Further, each sample was reciprocally labeled to allow us to quantify the effects of dye bias. Thus, each biological replicate was hybridized to four separate arrays, leading to 24 arrays for each antibody and 48 arrays in total. Finally, these 48 arrays were fully blocked across seven days, allowing us to robustly estimate the effect of date-of-hybridization (Fig. 1).

To quantify the importance of antibody selection we compared two polyclonal antibodies raised against the same N-terminal region of the Myc protein. One antibody was purchased from a commercial vendor while the other was locally produced (Fig. 1). Across the 48 arrays of our study, we observed a significant antibody effect. When the p-values of the top-ranked spots on the array are compared for the two antibodies, a linear trend is observed, suggesting specificity is comparable (Fig. 2A). However, despite equivalent replicate numbers and full-blocking of other parameters, as described below, p-values are distinctly lower for the commercial antibody, suggesting that it results in better signal to noise characteristics. To increase the specific activity of our home-made antibody, we subjected it to a purification procedure to remove contaminating serum proteins, and compared purified and unpurified home-made antibody in a similar fashion (Fig. 2B). A linear trend was again observed, with the purified antibody showing lower p-values than the unpurified. Note that the magnitude of p-values in Figures 2A and 2B are different because of the different replicate numbers involved (Fig. 2A, n = 24; 2B,

n=4 for each antibody). These data suggest that antibody purity is a key factor influencing the specificity and sensitivity of ChIP-chip studies, therefore antibody purification may improve the performance of some antibodies.

To evaluate the effect of array batch, dye bias and inter-day hybridization, we fit spot-wise linear models to the results of each parameter separately and plotted the Gaussian densities of the magnitudes of these effects (Fig. 3). The curves in the top portion of the figure show that for all three sources of variability (dye, batch and date) the majority of spots on the array show little to no effect. Most spots exhibit a fold change in the range of -0.5 to +0.5 $\log_2$ units, corresponding to an effect of at most 1.4-fold. Overall, no more than 10% of spots on the array show large variability towards any of these parameters – suggesting that both the ChIP protocol and the CpG island arrays used are robust towards perturbations in any of these factors.

Despite this basal similarity, though, there are differences across the three parameters. The curve of dye-bias shows significant asymmetry, corresponding to a consistent Cy3 bias. While only 11 spots show Cy5/Cy3 ratios of 1.4-fold or higher, 685 spots show Cy3/Cy5 ratios of 1.4-fold or greater. This marked asymmetry highlights the importance of balancing experiments across the two dyes used in the study.

The effect of array batch shows a similar, but smaller asymmetry. Older arrays generally show lower signal intensities than newer arrays, with 408 spots showing an old/new ratio of at least 1.4-fold, while only 210 spots show a new/old ratio of similar magnitude. This variability suggests that experiments should be conducted with arrays of a single batch, or experimental samples randomized across multiple batches.

Surprisingly, the effects of hybridization date affect about 10% of the spots on the

array, despite the same, experienced hands conducting all hybridizations in this series of experiments. This effect is symmetrical, with approximately equal numbers of spots showing date-dependent enrichment and depletion. The source of this bias is unknown. It is possible that atmospheric ozone may have differed from day-to-day. We and others have experienced Cy5 squelching due to ozone [32, 33]. However, in this case ozone is not likely to be the source of the bias, as these experiments were conducted during times of consistent low atmospheric ozone. It is also possible that variation in handling at the level of post-hybridization washes may contribute to this inter-day variability. Another possibility is that scanner calibration varied across the course of our experiment, but the biases described here are not temporally-dependent (data not shown). This suggests that if scanner-calibration is an issue, then it is a random drift rather than a progressive degradation.

Amplification method

Sufficient DNA for microarray hybridization cannot usually be obtained from a single ChIP reaction. The amount required can be achieved by pooling multiple ChIP reactions [12], but this approach is expensive and labor intensive. Further, when dealing with biologically limited samples, such as primary tumor material, repeated ChIPs may not be possible. Indeed, in such cases repeated ChIPs may introduce intra-sample heterogeneity as an additional confounding variable [34]. To overcome this rate-limiting step, the precipitated DNA in a single ChIP reaction is usually amplified. A variety of PCR-based methods exists, including ligation-mediated PCR (LM-PCR), random priming PCR (RP-PCR) and T7 priming PCR methods. Here we develop an assay to evaluate the

bias introduced by a given amplification protocol. We test the two most common amplification techniques employed in ChIP-chip experiments, RP-PCR and LM-PCR (Fig. 4A) and for the LM-PCR method we evaluate protocols with either low (LM-PCR A) or high (LM-PCR B) cycle number. Total genomic DNA was amplified with one of the three protocols, then hybridized against an equal amount of unamplified total genomic DNA (Fig. 4A). As shown in Figure 4B, for any given p-value threshold, LM-PCR A shows a smaller number of differentially bound spots than either RP-PCR or LM-PCR B. Thus, we show that it is possible to evaluate a variety of amplification protocols, using genomic DNA to identify which method minimizes bias.

Hybridization Control

A variety of different hybridization controls have been used for ChIP-chip studies. For example, some groups have employed a "no antibody" control [12], while others have used total genomic DNA [11, 35]. To determine if there are significant differences between these hybridization controls, we designed a series of experiments to test each approach (Fig. 5A). Specifically, we compared direct hybridization of ChIP samples against a mock treated control (Direct No Ab) and against an IgG antibody control (Direct IgG). We also considered the use of total input DNA as a "denominator" by hybridizing ChIP and IgG samples separately against total-input DNA (Indirect No Ab), then comparing their intensities indirectly during statistical analysis by using a linear model. The results (Fig. 5B) are surprisingly invariant from one control to the next, and at any given p-value the number of hits identified in each experiment is nearly identical. Nevertheless, the two direct designs show more sensitivity than the indirect approach.

This strongly suggests that direct controls are superior to the indirect approach, despite the loss of the "denominator effect". Another feature that adds merit to the direct design method, is that half as many arrays are required to analyze each biological replicate, compared to the indirect approach, thereby significantly decreasing the cost of ChIP-chip experiments.

c-Myc ChIP-chip parameter choice

We conducted a Myc ChIP-chip experiment using the optimized parameters determined above. Thus, we used a purified antibody for the ChIP reaction, DNA was amplified by LM-PCR, we hybridized a single CpG island array batch from a recent print run, and IgG was the choice as a hybridization control. Dye swaps were randomized across hybridization dates, and the entire experiment was conducted at high replication (n=13). By using this highly optimized design we expected to both minimize bias and increase specificity. Following pre-processing and statistical analysis of the raw data we obtained a list of ~535 Myc bound spots (Supplementary Table 1). Importantly, 35 of these 535 spots correspond to known Myc targets annotated in the Myc target gene database [36, 37]. To validate some of the novel target genes we performed Q-PCR from 4 conventional Myc ChIP biological replicates performed in HL60 cells. Importantly, these reactions used biologically independent samples that were completely separate from any of the samples used in the optimization studies or optimized array work. As a positive control we show that Myc is bound to the Cyclin D2 promoter and as a negative control that it is not bound to the chromosome 21 E-Box (Fig. 6). We designed the PCR primers against the promoter regions of five spots taken from the top 150 spots, ranked

by p-value, on the list. Four of these five (*GDAP2, RPL34, PHLPP* and *TCERG1*) were confirmed as being bound by Myc in growing HL60 cells (Fig. 6). Following the same criteria, we designed primers against the promoter regions of four spots taken from the middle third of the list, and all four (*CDC14A, CDCA2* and *NOL6*) were confirmed as being bound by Myc (Fig. 6). Finally, we designed primers against the promoter regions of four spots from the bottom third of the list, and three of these four (*PLTP, KLC4* and *GAS1*) were bound by Myc (Fig. 6). Thus, with our well established ChIP-chip experimental parameters we efficiently identified multiple novel *bona fide* Myc target genes. Importantly, high validation rates were observed across a wide range of p-values, suggesting that our analysis is conservative, and that additional Myc targets could be identified by relaxing the p-value threshold without significantly compromising specificity.

DISCUSSION

ChIP-chip is a relatively new technological advance that has quickly become the gold standard for screening protein-DNA interactions *in vivo*. However, it is also a highly complex method, with many tunable parameters. We set out to systematically test a large subset of these experimental design parameters. To do so, we employed a well-characterized system involving the Myc oncogene as a regulator of gene transcription in the HL60 human myeloid leukemia cells [12]. While this means that our results are formally restricted to a single biological system, the general nature of the parameters evaluated indicates a much broader relevance to ChIP-chip studies across many systems.

The first experimental design parameter considered was the influence of different antibodies on the results of a ChIP-chip study. Intuitively, the nature of the antibody is a limiting step for ChIP-chip studies. To put a lower-limit on antibody-specific variability, we selected two polyclonal antibodies raised against the same portion of the Myc protein (the N-terminal 262 residues, N262) that were derived from different sources. Critically, the two antibodies showed a broadly linear, well-ordered trend ($p < 2.2 \times 10^{-16}$; Wilcoxon Rank Test). However, the significance of differential hybridization varied substantially, with the commercially-available antibody being much more sensitive (Fig. 2A). We then subjected the locally-generated N262 antibody to a column purification procedure to remove contaminating serum proteins. The performance of this purified antibody is significantly improved compared to the unpurified antibody (Fig. 2B). Clearly, antibody selection is a critical element of ChIP-chip studies, and antibody specific activity is an important parameter to consider prior to hybridization.

Next we considered the effects of inter-batch variability of the CpG island

microarrays. This array platform has been widely used by other groups [26, 35, 38], giving confidence in its reliability. We tested the variability between two batches of arrays in a multi-factorial experiment. A consistent signal-intensity bias was observed, with the older array batch showing reduced signal relative to the newer batch (Fig. 3). This bias is highly reproducible, as it was evident in an analysis of 24 arrays from each batch. Bias of this magnitude is sufficient to create both false positives and false-negatives. Fortunately, there are straight-forward methods for controlling this factor. First, whenever possible, arrays from a single batch should be used. Second, when multiple batches are used, samples should be randomized or blocked across batches. Third, if linear-modeling methods are employed, explicit batch terms should be included in model-specification to remove this variability. The first two methods are standard experimental-design practices, while the third has been successfully used in the analysis of mRNA expression array data [39].

Our third question involves the effects of reciprocal-labeling (dye-swapping) a single sample with the Cy3 and Cy5 dyes. This topic has been extensively studied from both theoretical [40, 41], and empirical [42] perspectives. Even if reciprocal labeling has been shown to have less experimental value than true biological replicates, they appear to carry more information than simple repeated measures [42]. To our surprise, very moderate effects of reciprocal labeling were observed, with only 16 out of the >12,000 spots on the array showing a two-fold change in signal intensity based on dye-status (Fig. 3). This indicates that, at least for ChIP-chip with our CpG island arrays, reciprocal labeling is not necessary and dye-bias is a negligible confounder. Nevertheless, it seems prudent to include dye-status as a randomization factor in experimental design in the case

of unanticipated confounding dye-bias in specific experimental scenarios, especially because of the profound asymmetry in this parameter. Instead of dye-swapping a single sample we recommend simply balancing dye labels across experimental groups.

The fourth experimental parameter considered was the effect of hybridization-date. Because our experiment was saturated and fully blocked across all factors, we were able to directly estimate the effects of this parameter with a linear model. We were surprised to find that inter-day hybridization had more influence than either batch-effect or dye-swap status (Fig. 3). The sources of this bias are unclear. The symmetrical distribution of the effect suggests that neither ozone-status nor scanner settings are a major factor. Perhaps use of a mechanical hybridization and wash station would decrease inter-day variability. Control of this factor is straight-forward: samples should be randomized across experimental day, so that biological replicates of a given condition are hybridized on different days.

The fifth experimental parameter considered was the method of DNA amplification. While the antibody-based ChIP protocol imparts great specificity and sensitivity to ChIP-chip experiments, most ChIP protocols result in a relatively small amount of DNA. Because the quantity of DNA necessary for hybridization cannot usually be obtained in a single ChIP, two basic strategies have evolved: pooling of multiple ChIPs [12] and amplification of ChIP DNA [25]. A variety of different PCR-based amplification methods have been developed and are in use by different groups. Assessing the performance of these methods has been a challenging problem. To directly compare an amplified ChIP sample to an unamplified ChIP sample requires the pooling of a large number of ChIPs for each replicate, resulting in significant costs, time and

labor. Further, care must be taken in the experimental design to ensure that the reduction in variance that results from pooling ChIP reactions does not confound the effects of the amplification procedure. As a result of these financial and technical factors, we are not aware of any detailed, replicate studies measuring the performance of amplification methods for ChIP-chip data, although such studies do exist for mRNA experiments [43, 44], where a much larger amount of starting material is readily available. To resolve this issue, we have developed a novel approach that hybridizes amplified against non amplified genomic DNA (Fig. 4A). We evaluated three amplification protocols: RP-PCR, or one of the two protocols of LM-PCR. The matched amplified and unamplified sample pairs were then hybridized to 12k CpG island microarrays in equal quantities. Because of its simplicity, this procedure could be readily repeated and an n=4 was generated for each amplification procedure. The number of spots found to be differentially hybridized between the matched amplified and unamplified samples is then taken as a measure of amplification bias, and is plotted for a range of p-values (Fig. 4B). Interestingly, the performance of random-priming falls in between that of the two LM-PCR protocols, indicating that specific parameter choices within a protocol are important for determining overall performance. Our method of comparing amplification procedures provides a new way to measure variability, independent of antibody or biological question, at a reasonable cost.

The final experimental design parameter considered was the selection of a negative control for hybridization to the array. We considered three approaches (Fig. 5A), including two direct and one indirect method. For each method we performed four Myc ChIP-chip experiments in HL60 cells. When we compared the sensitivity of these

methods (Fig. 5B) the two direct methods performed comparably, with no significant differences between the IgG and no-antibody controls. The indirect design, however, did show increased variability, with reduced sensitivity at any given p-value. Thus, direct designs, which use half as many arrays as indirect designs, are favored for both cost and sensitivity reasons.

Finally, we set up a Myc ChIP-chip experiment to test these optimized parameters in an integrated way. Following extensive biological replication (n=13) we validated putative Myc target genes by Q-PCR from conventional ChIP reactions. Importantly, our validation used biologically independent samples to ensure maximal generalization of the results. We validated 4 of 5 genes from the top third of the list (*GDAP2, RPL34, PHLPP* and *TCERG1*), the 4 of 4 genes taken from the middle third of the list (*CDC14A, CDCA2* and *NOL6*), and 3 of 4 genes taken from the bottom third of the list (*PLTP, KLC4* and *GAS1*) (Fig. 6). One reason of getting false-positives in the ChIP-chip experiment could be the non-specific DNA-binding property of Myc [45]. This feature is most likely the reason of why some targets are identified as potential Myc-binding sites, but are not validated upon further testing. In total, then, we have identified 535 novel putative Myc target genes, of which 11 have been validated by gold-standard Q-PCR. In addition, 35 spots out of the 535 spots represented genes whose expression is known to be regulated by Myc. Moreover, 85 spots corresponded to genes that belong to a family in which another member has been described as a Myc-regulated gene [37]. Five of these have been tested and validated (*GDAP2, RPL34, CDC14A, CDCA2* and *NOL6*) suggesting the remaining 80 targets are likely to also be Myc target genes. Importantly, we showed that the validation rate remained high in the bottom portion of our list. This suggests that the

statistical analysis was conservative, and that additional hits might be uncovered by relaxing our p-value threshold from 0.05 to 0.10 or higher. This is particularly important because many published ChIP-chip studies use a p-value threshold of 0.001, 50 times more stringent than used here [11]. Overall, validation of the Myc target genes identified in this study confirms the success of our parameter optimizations.

Over the past decade, a series of studies from multiple groups served to characterize the major experimental design parameters for mRNA expression studies [28, 42, 46-48]. A common theme of these studies is that it is necessary to use both empirical and statistical methodologies to optimize the experimental design. We have provided the first comprehensive overview of experimental design parameters for ChIP-chip data, testing both parameters related to all microarray experiments and those unique to ChIP-chip studies. We find good concordance of ChIP-chip and mRNA expression design characteristics. For example, in both cases, indirect hybridization designs appear to be more costly without providing a significant advantage over simpler direct designs. Similarly, we have shown that biological replicates are more important than dye-swaps, just as identified for mRNA expression arrays [42]. We also demonstrated that antibody characteristics and amplification methods are major sources of variability in ChIP-chip studies. For example, using an antibody of maximal specific activity appears to be critical for increasing the sensitivity of ChIP-chip studies. In addition, we introduced a novel method for testing amplification bias in an antibody-independent manner. This study indicated that amplification procedures can significantly bias ChIP-chip results, and provides a simple, efficient method for benchmarking the bias of specific amplification protocols.

Our findings also raise a number of intriguing questions surrounding the antibody and amplification results. We demonstrate that antibody-specific baseline variability exists in ChIP-chip studies, and that while it can be mitigated it cannot be completely removed. It would be of great interest to replicate these studies with monoclonal antibodies, or polyclonal antibodies raised against different regions of a protein. In this manner, it might be possible to identify characteristics that predict which antibodies will be most sensitive or specific in a ChIP-chip study. Alternatively, this may not be a generalizable parameter and may be protein-specific. Similarly, we have presented a novel technique for studying amplification bias, and identified a variant of ligation-mediated PCR as most favorable in this regard. Nevertheless, all amplification protocols tested here introduced a significant amount of bias. Our strategy for bias-identification should prove useful in the development of reduced-bias amplification protocols.

As ChIP-chip studies continue to broaden in the scope and detail of the biological questions they probe, the application of robust and validated experimental designs will be a significant advantage to the field.

CONCLUSIONS

We performed a series of empirical studies to systematically evaluate the effects of key design parameters for ChIP-chip studies. Antibody purity is an important factor for controlling the specificity and sensitivity of ChIP-chip experiments, thus, antibody purification should be routinely considered to improve assay performance. The use of multiple array batches increased bias, with older arrays showing lower signal intensity than newer arrays, indicating that either a single batch of arrays be used, or experimental samples be randomized across batches. Minimal dye bias was observed in our study, although prudent experimental design would include dye swap as a randomization factor by balancing dye labels across experimental groups to guard against unexpected biases. The date of hybridization was a major source of variability, but the source of this factor remains unclear. Evidently, samples should be randomized across experimental days. Three DNA amplification methods were evaluated by a novel approach that hybridizes amplified against non-amplified genomic DNA. This is a simple way to evaluate the bias of an amplification protocol, independent of the biological question. We found that ligation-mediated PCR introduced less bias than random-priming PCR, although bias was found to be sensitivity to parameterization. Finally, we found that direct hybridization controls reduced variability and increased sensitivity relative to indirect controls. Thus, direct designs are preferential for both sensitivity and financial reasons. To integrate these findings, we tested the optimized parameters in a Myc ChIP-chip experiment in HL60 cells. High validation rates observed by Q-PCR in independent samples show the sensitivity and specificity of our optimized assay. This systematic and thorough optimization study is a key step towards putting our understanding of ChIP-chip

experimental parameters on an equal footing with our understanding of mRNA expression arrays, and will enable improved elucidation of gene-regulatory networks.

MATERIALS AND METHODS

Antibody production and purification

The DNA fragment corresponding to the Myc 1-262 N-terminal domain polypeptide was cloned into pET15b vector (Novagen 69661-3) at 5'-NdeI and 3'-BamHI sites. His-c-Myc (1-262) fusion protein was purified under denatured conditions using Talon beads (BD Biosciences) according to the manufacturer's protocol. The purified protein was concentrated to 2.8 mg/ml before injecting into the rabbit. The purified protein was then dissolved in a volume of 500 µL of PBS per rabbit. Primary immunizations were done with 500 µg of purified protein per rabbit. First boost was done with 250 µg of purified protein. Two subsequent boosts were done with 50 µg of purified protein. The final bleed was then purified using Protein A, IgG Purification (Pall PN5300-IgG proA) according to manufacturer's protocol.

Cell Culture conditions

HL60 (ATCC) cells were maintained in α-MEM with 10% FBS.

Chromatin immunoprecipitation

Exponentially growing HL60 cells were cross-linked with 1% formaldehyde for 10 minutes at 37 °C. The crosslinking reaction was quenched by addition of glycine to a final concentration of 0.125 M for 5 minutes, followed by two washes with phosphate-buffered saline (PBS). Cells were resuspended in cell lysis buffer (5 mM PIPES pH 8, 85 mM KCl, 0.5% [v/v] NP40, 1 mM PMSF, 10 µg/mL aprotinin, 10 µg/mL leupeptin) for 10 minutes on ice and then pelleted (5000 rpm, 5 minutes, 4 °C). The pellet was resuspended in 1 mL of nuclei lysis buffer (50mM Tris-HCl pH 8.1, 10mM EDTA, 1% SDS, 1 mM PMSF, 10 µg/mL aprotinin, 10 µg/mL leupeptin) for 10 minutes on ice and then sonicated using 8 pulses (12-13 Watts, setting 10, 10 seconds per pulse, 45 seconds on ice between pulses) from a Model 60 Sonic Dismembrator (Fisher Scientific 15-338-53) to generate fragments between 600 bp and 1000 bp. Lysates were centrifuged for 10 minutes at 21,000 x g at 4 °C. Supernatants were diluted into an equal volume with IP dilution buffer (0.01% SDS, 1.1% Triton-X100, 1.2mN EDTA, 16.7mM Tris-HCl pH 8.1, 0.2% Sarkosyl, 1 mM PMSF, 10 µg/mL aprotinin, 10 µg/mL leupeptin) and pre-cleared for 30 minutes at 4 °C with protein G-PLUS agarose beads (Santa Cruz Biotechnology sc-2002). Prior to use, G-PLUS agarose beads were blocked with salmon

sperm DNA at a final concentration of 50 µg/mL and rotated overnight at 4 °C. Diluted and cleared extracts corresponding to 10 x 10^6 HL60 cells were incubated and rotated at 4°C for approximately 12 to 16 hours with each of the following antibodies: no-antibody control, 0.7 µg normal rabbit IgG (Santa Cruz Biotechnology sc-2027), 0.7 µg N262 (Santa Cruz Biotechnology sc-764), for the N262 home-made unpurified and purified antibodies, we determined empirically, by serial dilutions, the amount of antibody to be used. 50 µL of salmon sperm DNA pre-blocked Protein G-PLUS agarose beads were added to each sample, incubated on a rotating platform at 4 °C for 3 hours. Each pellet was washed once with 1.4 mL of sonication buffer and then twice with 1.4 mL of high salt buffer (0.1% [v/v] SDS, 1% [v/v] Triton X-100, 1 mM EDTA, 50 mM HEPES, 500 mM NaCl, 0.1% [w/v] sodium deoxycholate) and then once with 1.4 mL LiCl Buffer (250 mM LiCl, 1% [v/v] NP-40, 1% [w/v] sodium deoxycholate, 1 mM EDTA, 1 mM Tris pH 8) and finally twice with 1.4 mL TE pH 8 (10 mM Tris pH8, 1 mM EDTA). For each wash, the pellets were mixed for 5 minutes at room temperature then pelleted (3000 rpm, 30 seconds, room temperature). After the last wash, the pellets were eluted in 300 µL of Elution buffer (1% [w/v] SDS, 10 mM Tris pH 8, 5 mM EDTA), incubated at 65 °C for 15 minutes, and then pelleted (3000 rpm, 3 minutes, room temperature). Cross-links were reversed in the presence of 200 mM NaCl at 65 °C over-night and samples were treated with RNase A (Sigma R5500). After ethanol precipitation, the samples were resuspended in 100 µL of TE (10 mM Tris, pH 7.5, 1 mM EDTA), 25 µL of 5x proteinase K buffer (1.25% SDS, 50 mM Tris, pH 7.5, 25 mM EDTA), and 1.5 µL of proteinase K (Roche 1413783) and incubated at 42 °C for 2 hours. DNA was extracted from each sample using phenol:chloroform:isoamyl alcohol (25:24:1), then precipitated with 1/10th volume of 3 M sodium acetate (pH 5.3), 5 µg of glycogen, and 2 volumes of ethanol at -20 °C overnight. Pellets were collected by microcentrifugation and resuspended in 60 µL of H_2O.

PCR

A total of 2 µL of the ChIP DNA was PCR amplified using promoter-specific primers. All PCR primers (synthesized by Sigma Genosys or Invitrogen) were resuspended in sterile RNase-/DNase-free water. In general, each 10 µL PCR reaction, set up at room temperature, was composed of 10 ng/µL of each primer (fwd/rev), 1x PCR buffer, 0.2

mM dNTP mix, sterile water, DNA template, and 0.2 µL HotstarTaq DNA polymerase (Qiagen 203205); the HotstarTaq was activated by initiating the PCR program at 95°C for 15 minutes. Amplification of each PCR primer set was performed using multiple annealing temperatures along a 52 °C to 68 °C gradient for 30 seconds. Primer sequences are provided in Supplementary Table 2. All PCR reactions were electrophoresed on 100 mL 1.2% agarose gels containing 4 µL of 10 mg/mL EtBr, and then visualized by UV florescence.

Real-time quantitative PCR (Q-PCR)

Real-time quantitative PCR amplification was conducted using the SYBR Green assay in the ABI PRISM 7900-HT (Applied Biosystems). Each 12 µL quantitative PCR reaction was composed of the following: 2 µL of the DNA isolated in the ChIP, 1x PCR Buffer, 2.5 mM $MgCl_2$, 0.17 mM dNTP mix, sterile water, 0.25x SYBR Green (Sigma S9430), 0.2 µL Rox reference dye (Invitrogen 12223-012) and 0.05 µL Platinum Taq DNA polymerase (Invitrogen 10966-034) and was performed in triplicate in a 384-well plate. The reactions began at 50 °C for 2 minutes and then were activated at 95 °C for 10 minutes followed by 40 cycles of 95 °C for 15 seconds and 60 °C for 1 minute. Human male genomic DNA (Novagen 70572) was used as standard real-time quantitative PCR (Q-PCR) was conducted in triplicates for each of the four independent biological replicates. The Q-PCR data was analyzed by calculating the Myc/IgG ratios for each target gene, and the Chr21 negative control. A symmetrical distribution was then obtained by log2-transformation of each individual ratio. Statistical testing involved comparison of the ratios for each target gene to the ratios of the Chr21 negative control using a two-tailed paired t-test. Primer sequences are provided in Supplementary Table 3. The primers were designed using Primer Express Software Version 2.0 according to the manufacturer procedures (Applied Biosciences).

Random primer PCR amplification method

The immunoprecipitated DNA was amplified as described elsewhere [49]with the following modifications: 15 µL of Round A are mixed with Round B1 Reaction Mix to a final volume of 100 µL [4µL of 50 mM $MgCl_2$, 10 µL of 10x PCR Buffer, 2 µL of 10mM dNTPs, 2.5 µL of 100 µM of Primer B (5'GTTTCCCAGTCACGATC3'), 1 µL of Taq DNA Polymerase (Promega M1861)]. Amplification/Nucleotide Incorporation Program:

(94 °C for 30 seconds, 40 °C for 30 seconds, 50 °C for 30 seconds, 72 °C for 1 minute) x 20 cycles, 72 °C for 2 minutes.

15 µL of Round B1 were mixed with Round B2 Reaction Mix to a final volume of 50 µL [2 µL of 50 mM $MgCl_2$, 5 µL of 10X PCR Buffer, 0.5 µl of dNTP Mix B2 (25mM of dATP, dCTP, dCTP, 12.5 mM dTTP), 1.5 µL of 2 mM aa-dUTP, 1 µL of 100 µM of Primer B (5'GTTTCCCAGTCACGATC3'), 1 µL of Taq DNA Polymerase (Promega M1861)]. Amplification/Nucleotide Incorporation Program: (92 °C for 30 seconds, 40 °C for 30 seconds, 50 °C for 30 seconds, 72 °C for 1 minute) x 15 cycles, 72 °C for 2 minutes. Round B2 amplified DNA was concentrated using Microcon columns (Millipore, Microcon YM-30). Once the ChIP DNA was amplified to 3 µg, PCR reactions were performed with 10 ng amplified DNA and were identical to the reactions performed after the ChIP DNA was initially isolated. This PCR was used to determine whether specificity between the IgG or NoAb and antibody samples was maintained after the amplification.

Ligation-mediated PCR (LM-PCR) Amplification

The DNA from each ChIP reaction was treated as described elsewhere [25]. The PCR was performed using the following cycling protocol: 1 cycle of 55 °C for 2 minutes, 72 °C for 5 minutes, and 95 °C for 2 minutes and the corresponding number of cycles for LM-PCR A and LM-PCR B (see below LM-PCR A and B cycle conditions) at 95 °C for 30 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute, ending with 4 minutes at 72 °C and then an endless cooling at 4 °C. After the PCR, the DNA was again isolated using the QIAquick PCR purification kit (Qiagen 28106). The concentration of the purified DNA was determined on the NanoDrop ND-1000 spectrophotometer. Often the 3 µg required for hybridization was not acquired after this first round of amplification, so another round of amplification would be performed, followed by purification and determination of the DNA concentration as previously described. LM-PCR A amplification: 20 cycles followed by QIAquick PCR purification, and then 15 cycles followed by QIAquick PCR purification. LM-PCR B amplification: 20 cycles followed by QIAquick PCR purification, repeated 3 rounds, sequentially. Once the ChIP DNA was amplified to 3 µg, PCR reactions of positive and negative control genes (Supplementary Table 2) were performed with 10 ng amplified DNA and were identical to the reactions performed after

the ChIP DNA was initially isolated. This PCR was necessary to determine whether specificity between the IgG or NoAb and antibody samples was maintained after the amplification.

Labeling reaction

A total of 3 µg of amplified DNA from the antibody sample, no antibody, or IgG sample were vacuum desiccated and resuspended in 2.5 µL of H_2O (Sigma W4502). 4.5 µL of 0.1 M $NaHCO_3$ at pH 9 was added. The DNA was resuspended by vortexing several times. 2 µL of Cy3 or Cy5 dye (Amersham, PA23001 and PA25001 respectively) were added to the sample. The mix was incubated for 1.5 hours at room temperature in the dark. After incubation, 35 µL of 100 mM Na-Acetate pH5.2 was added and topped with H_2O (Sigma W4502) to a total volume of 100 µL for each sample. The fluorescent labeled probes were purified using QIAquick PCR Purification kit (Qiagen 28106). The DNA was eluted with 50 µL of EB buffer from the kit, heated to 65°C. The 100 µL combined ChIP and no antibody samples were vacuum desiccated using a SpeedVac for 60 minutes at maximum heat.

DNA Microarray Hybridization

The mixed DNA was resuspended in 5 µL of H_2O and mixed with 85 µL hybridization mix (100 µL of DIG Easy Hyb solution (Roche 603 558), 5 µL of 10 mg/mL calf thymus DNA (Sigma D8661), 5 µL of 10 mg/mL yeast tRNA (Invitrogen 15401-029)). The resulting solution was incubated at 65 °C for 2 minutes, then cooled to room temperature and applied to a CpG island array slide (UHN Microarray Centre HCGI12K) and incubated overnight at 42 °C in a hybridization chamber.

Microarray washes

After hybridization, the microarray slide was washed once with washing buffer 1 (1x SSC, 0.1% SDS) at 50°C for 5 minutes. This was followed by a wash with washing buffer 1 (room temperature for 5 minutes) and 2 washes with washing buffer 2 (0.2x SSC for 5 minutes) and one final wash with washing buffer 3 (0.1x SSC for 5 minutes at room temperature). The slide was then dried by spinning at 700 rpm for 15 minutes. To preserve the Cy5 dye from ozone-dependent degradation, which was observed during the summertime when the hybridization occurred, the slides were then dipped in DyeSaver 2 (Genisphere Q500600) and allowed to air dry.

Microarray Scanning

The microarray slides were scanned using a Gene Pix 4000B scanner (Axon Instruments) at multiple PMT voltages.

Microarray Data Pre-Processing

Array images were first manually examined for image artifacts, and then quantitated using GenePix Pro (v6.0.1.27). A scan that maximized the dynamic range without saturating high-intensity signals was selected and carried forward for analysis. This scan was quantified using the "circular" segmentation algorithm, then pre-processed using the variance-stabilizing normalization algorithm (VSN). VSN was implemented in the R statistical environment (v2.4.1) in the BioConductor open-source project [50]. Version 1.12.0 of the vsn package was employed, with print-tip groups used as strata, and default parameterization except for using 1000 iterations for fitting (increased from the default of 10). All raw and processed array data will be made publicly available in the Gene Expression Omnibus (GEO) repository at NCBI – an accession will be available prior to publication.

Significance Testing

To test for differential hybridization, we exploited general linear modeling, as implemented in the limma package (v2.9.13) [51] of the BioConductor open-source library [50] for the R statistical environment (v2.4.1). First, a spot-wise linear model was fit separately to each major design parameters: array batch, hybridization date, and dye-status. No interaction effects were included due to lack of physical motivation. A Bayesian moderation of standard error was employed following model-fitting [51]. A Gaussian kernel density was then employed on the effect estimates to compare the relative magnitudes of these design parameters. Next, a spot-wise linear model was fit comparing the hybridizations using a commercially available antibody to those using a locally produced one. The top 100 spots from each analysis were identified, and the pair-wise p-values plotted to identify trends in antibody specificity. A similar analysis was performed for purified and non-purified locally-produced antibody hybridizations. All model-fitting again employed the limma package (v2.9.13). To compare different amplification methods we fit separate linear models to the amplified vs. non-amplified comparisons for each amplification method. The number of differentially bound spots

between the two samples was calculated across the entire range of naïve p-value thresholds from 0 to 1, at 0.00001 unit resolution. A similar analysis was used to investigate hybridization order, except that the models were fit simultaneously and the effect of each hybridization order was extracted separately using a contrast matrix. An identical tiling across p-values was used, in this case in the range 0 to 0.10 with a resolution of 0.0001. These analyses were performed in R (v2.4.1). All gene-lists presented are based on a p-value threshold of $p < 0.05$.

Spot Annotation

The published array annotations for the 12k UHN CpG island microarray were based on the UCSC build hg17 of the human genome [30]. To update these results for the more polished hg18 build we implemented a new annotation algorithm using a BioPerl (v1.5.1) based Perl (v5.8.8) script [52]. First, the 3' and 5' sequence-reads were overlapped using bl2seq [53] with a word-size of 7. In cases where a minimum 50 base-pair overlap was found the aggregated sequence was selected; otherwise the longer of the two sequence-reads was chosen. Next, the sequences were aligned to the RepeatMasker filtered chromosomal sequence-reads of human genome build hg18 [54] using a locally-compiled (gcc v4.1.1; AIX 5.2.0.0) NCBI BLASTN executable with a word-size of 7 [55]. The best matching hit was selected as the most likely region of hybridization, and the ratio of the lengths of the top two hits was used as a measure of the likelihood of cross-hybridization.

LIST OF ABBREVIATIONS USED

ChIP-chip: Combination of Chromatin Immunoprecipitation with high-throughput DNA microarray technology

PCR: Polymerase chain reaction

Q-PCR: Real time quantitative-PCR

LM-PCR: Ligation mediated-PCR

RP-PCR: Random primer-PCR

Ab: Antibody

NoAb: No antibody

IgG: Immunoglobuline G

AUTHORS' CONTRIBUTIONS

R.P., P.C.B. and L.Z.P. designed and initiated the project. R.P. performed the majority of the experiments, including the target gene validations of Fig. 6. P.C.B. performed all bioinformatics analysis. S.K., A.S., A.P.H. and F.K. performed the ChIP-chip reactions analyzed in Fig. 6. All authors edited, reviewed, and approved the final manuscript.

ACKNOWLEDGEMENTS

The authors acknowledge support from the U.S. Department of Defense Breast Cancer Research Program grant BC032138 (LZP), a Knudson Postdoctoral Fellowship (RP), fellowships from the PreCarn Foundation (PCB), the Natural Sciences and Engineering Research Council (PCB), and the CIHR EIRR21st STIHR (PCB), the Terry Fox Foundation Fellowship through an award from the National Cancer Institute of Canada (SK), Ontario Graduate Scholarship (AS, APH). We also thank the Penn Lab for helpful comments during the course of these experiments and in reviewing the manuscript, and to Mr. Richard Lu for computer system support and administration.

REFERENCES

1. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO: **Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF**. *Nature* 2001, **409**(6819):533-538.
2. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I *et al*: **Transcriptional regulatory networks in Saccharomyces cerevisiae**. *Science* 2002, **298**(5594):799-804.
3. Lieb JD, Liu X, Botstein D, Brown PO: **Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association**. *Nat Genet* 2001, **28**(4):327-334.
4. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E *et al*: **Genome-wide location and function of DNA binding proteins**. *Science* 2000, **290**(5500):2306-2309.
5. Bieda M, Xu X, Singer MA, Green R, Farnham PJ: **Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome**. *Genome Res* 2006, **16**(5):595-605.
6. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR *et al*: **Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1**. *Cell* 2005, **122**(1):33-43.
7. Cheng AS, Jin VX, Fan M, Smith LT, Liyanarachchi S, Yan PS, Leu YW, Chan MW, Plass C, Nephew KP *et al*: **Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters**. *Mol Cell* 2006, **21**(3):393-404.
8. Fernandez PC, Frank SR, Wang L, Schroeder M, Liu S, Greene J, Cocito A, Amati B: **Genomic targets of the human c-Myc protein**. *Genes Dev* 2003, **17**(9):1115-1129.
9. Horak CE, Mahajan MC, Luscombe NM, Gerstein M, Weissman SM, Snyder M: **GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis**. *Proc Natl Acad Sci U S A* 2002, **99**(5):2924-2929.
10. Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, Green R, Farnham PJ: **Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27**. *Genes Dev* 2004, **18**(13):1592-1605.
11. Li Z, Van Calcar S, Qu C, Cavenee WK, Zhang MQ, Ren B: **A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells**. *Proc Natl Acad Sci U S A* 2003, **100**(14):8164-8169.
12. Mao DY, Watson JD, Yan PS, Barsyte-Lovejoy D, Khosravi F, Wong WW, Farnham PJ, Huang TH, Penn LZ: **Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression**. *Curr Biol* 2003, **13**(10):882-886.

13. Pokholok DK, Zeitlinger J, Hannett NM, Reynolds DB, Young RA: **Activated signal transduction kinases frequently occupy target genes**. *Science* 2006, **313**(5786):533-536.
14. Squazzo SL, O'Geen H, Komashko VM, Krig SR, Jin VX, Jang SW, Margueron R, Reinberg D, Green R, Farnham PJ: **Suz12 binds to silenced regions of the genome in a cell-type-specific manner**. *Genome Res* 2006, **16**(7):890-900.
15. Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ: **Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis**. *Genes Dev* 2002, **16**(2):235-244.
16. Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA *et al*: **A tiling resolution DNA microarray with complete coverage of the human genome**. *Nat Genet* 2004, **36**(3):299-303.
17. Takayama K, Kaneshiro K, Tsutsumi S, Horie-Inoue K, Ikeda K, Urano T, Ijichi N, Ouchi Y, Shirahige K, Aburatani H *et al*: **Identification of novel androgen response genes in prostate cancer cells by coupling chromatin immunoprecipitation and genomic microarray analysis**. *Oncogene* 2007.
18. Guo QM, Malek RL, Kim S, Chiao C, He M, Ruffy M, Sanka K, Lee NH, Dang CV, Liu ET: **Identification of c-myc responsive genes using rat cDNA microarray**. *Cancer Res* 2000, **60**(21):5922-5928.
19. Kannan K, Kaminski N, Rechavi G, Jakob-Hirsch J, Amariglio N, Givol D: **DNA microarray analysis of genes involved in p53 mediated apoptosis: activation of Apaf-1**. *Oncogene* 2001, **20**(26):3449-3455.
20. O'Connell BC, Cheung AF, Simkevich CP, Tam W, Ren X, Mateyak MK, Sedivy JM: **A large scale genetic analysis of c-Myc-regulated gene expression patterns**. *J Biol Chem* 2003, **278**(14):12563-12573.
21. Stanelle J, Stiewe T, Theseling CC, Peter M, Putzer BM: **Gene expression changes in response to E2F1 activation**. *Nucleic Acids Res* 2002, **30**(8):1859-1867.
22. Watson JD, Oster SK, Shago M, Khosravi F, Penn LZ: **Identifying genes regulated in a Myc-dependent manner**. *J Biol Chem* 2002, **277**(40):36921-36930.
23. Oberley MJ, Farnham PJ: **Probing chromatin immunoprecipitates with CpG-island microarrays to identify genomic sites occupied by DNA-binding proteins**. *Methods Enzymol* 2003, **371**:577-596.
24. Ren B, Dynlacht BD: **Use of chromatin immunoprecipitation assays in genome-wide location analysis of mammalian transcription factors**. *Methods Enzymol* 2004, **376**:304-315.
25. Oberley MJ, Tsao J, Yau P, Farnham PJ: **High-throughput screening of chromatin immunoprecipitates using CpG-island microarrays**. *Methods Enzymol* 2004, **376**:315-334.
26. Paris J, Virtanen C, Lu Z, Takahashi M: **Identification of MEF2-regulated genes during muscle differentiation**. *Physiol Genomics* 2004, **20**(1):143-151.
27. Liu CL, Schreiber SL, Bernstein BE: **Development and validation of a T7 based linear amplification for genomic DNA**. *BMC Genomics* 2003, **4**(1):19.
28. Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY *et al*: **The MicroArray Quality Control**

(MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006, **24**(9):1151-1161.

29. Wei C, Li J, Bumgarner RE: **Sample size for detecting differentially expressed genes in microarray experiments**. *BMC Genomics* 2004, **5**(1):87.
30. Heisler LE, Torti D, Boutros PC, Watson J, Chan C, Winegarden N, Takahashi M, Yau P, Huang TH, Farnham PJ *et al*: **CpG Island microarray probe sequences derived from a physical library are representative of CpG Islands annotated on the human genome**. *Nucleic Acids Res* 2005, **33**(9):2952-2961.
31. Birnie GD: **The HL60 cell line: a model system for studying human myeloid cell differentiation**. *Br J Cancer Suppl* 1988, **9**:41-45.
32. Branham WS, Melvin CD, Han T, Desai VG, Moland CL, Scully AT, Fuscoe JC: **Elimination of laboratory ozone leads to a dramatic improvement in the reproducibility of microarray gene expression measurements**. *BMC Biotechnol* 2007, **7**:8.
33. Fare TL, Coffey EM, Dai H, He YD, Kessler DA, Kilian KA, Koch JE, LeProust E, Marton MJ, Meyer MR *et al*: **Effects of atmospheric ozone on microarray data quality**. *Anal Chem* 2003, **75**(17):4672-4675.
34. Bachtiary B, Boutros PC, Pintilie M, Shi W, Bastianutto C, Li JH, Schwock J, Zhang W, Penn LZ, Jurisica I *et al*: **Gene expression profiling in cervical cancer: an exploration of intratumor heterogeneity**. *Clin Cancer Res* 2006, **12**(19):5632-5640.
35. Vigano MA, Lamartine J, Testoni B, Merico D, Alotto D, Castagnoli C, Robert A, Candi E, Melino G, Gidrol X *et al*: **New p63 targets in keratinocytes identified by a genome-wide approach**. *Embo J* 2006.
36. **Myc target gene database**. In.: http://www.myc-cancer-gene.org/site/mycTargetDB.asp.
37. Zeller KI, Jegga AG, Aronow BJ, O'Donnell KA, Dang CV: **An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets**. *Genome Biol* 2003, **4**(10):R69.
38. Krieg AJ, Hammond EM, Giaccia AJ: **Functional analysis of p53 binding under differential stresses**. *Mol Cell Biol* 2006, **26**(19):7030-7045.
39. Semeralul MO, Boutros PC, Likhodi O, Okey AB, Van Tol HH, Wong AH: **Microarray analysis of the developing cortex**. *J Neurobiol* 2006.
40. Dobbin K, Shih JH, Simon R: **Statistical design of reverse dye microarrays**. *Bioinformatics* 2003, **19**(7):803-810.
41. Dobbin KK, Kawasaki ES, Petersen DW, Simon RM: **Characterizing dye bias in microarray experiments**. *Bioinformatics* 2005, **21**(10):2430-2437.
42. He YD, Dai H, Schadt EE, Cavet G, Edwards SW, Stepaniants SB, Duenwald S, Kleinhanz R, Jones AR, Shoemaker DD *et al*: **Microarray standard data set and figures of merit for comparing data processing methods and experiment designs**. *Bioinformatics* 2003, **19**(8):956-965.
43. Dafforn A, Chen P, Deng G, Herrler M, Iglehart D, Koritala S, Lato S, Pillarisetty S, Purohit R, Wang M *et al*: **Linear mRNA amplification from as little as 5 ng total RNA for global gene expression analysis**. *Biotechniques* 2004, **37**(5):854-857.
44. Marko NF, Frank B, Quackenbush J, Lee NH: **A robust method for the**

amplification of RNA in the sense orientation. *BMC Genomics* 2005, **6**(1):27.
45. Dang CV, van Dam H, Buckmire M, Lee WM: **DNA-binding domain of human c-Myc produced in Escherichia coli**. *Mol Cell Biol* 1989, **9**(6):2477-2486.
46. Pan W, Lin J, Le CT: **How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach**. *Genome Biol* 2002, **3**(5):research0022.
47. Peixoto BR, Vencio RZ, Egidio CM, Mota-Vieira L, Verjovski-Almeida S, Reis EM: **Evaluation of reference-based two-color methods for measurement of gene expression ratios using spotted cDNA microarrays**. *BMC Genomics* 2006, **7**:35.
48. Shih JH, Michalowska AM, Dobbin K, Ye Y, Qiu TH, Green JE: **Effects of pooling mRNA in microarray class comparisons**. *Bioinformatics* 2004, **20**(18):3318-3325.
49. **Immunoprecipitated DNA Amplification** In. Edited by http://www.microarrays.ca/support/PDFs/ChIP_Protocol_for_Microarray_Analysis_Beads_corr_01-12-06.pdf.
50. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J *et al*: **Bioconductor: open software development for computational biology and bioinformatics**. *Genome Biol* 2004, **5**(10):R80.
51. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments**. *Stat Appl Genet Mol Biol* 2004, **3**:Article3.
52. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H *et al*: **The Bioperl toolkit: Perl modules for the life sciences**. *Genome Res* 2002, **12**(10):1611-1618.
53. Tatusova TA, Madden TL: **BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences**. *FEMS Microbiol Lett* 1999, **174**(2):247-250.
54. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ *et al*: **The UCSC Genome Browser Database**. *Nucleic Acids Res* 2003, **31**(1):51-54.
55. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25**(17):3389-3402.

Figure Legends

Figure 1: Experimental design to evaluate antibody effect, array batch, reciprocal labeling and hybridization. Schematic of the experimental design in which six independent biological replicates of cross-linked HL60 cells were subjected to ChIP with either commercial or home-made antibodies raised against the N-terminal 262 residues of the Myc oncoprotein. Each sample was hybridized to individual arrays with its matched No Ab control. Dashed arrows indicate biological replicates #2 to #5, not included in the figure for simplicity. Black arrows indicate reciprocal labeling. Dotted lines indicate the period of seven days in which the 48 arrays were hybridized.

Figure 2: Measuring antibody effect

(A) Comparison of ChIP-chip results using a commercial and home-made polyclonal antibody raised against the same portion of Myc protein showed a similar linear trend and ordering by the two antibodies ($p < 2.2 \times 10^{-16}$; Wilcoxon Rank Test), however, they differed by orders of magnitude in sensitivity (axis scales). (B) Purification of the home-made antibody preparation improved sensitivity. The line represents equal sensitivity ($y = x$). Note that orders of magnitude on the axes differ between parts A and B because the replicate numbers are different (A, $n = 24$; B, $n = 4$).

Figure 3: Evaluating array batch effects, dye-bias, and inter-day processing

To assess the effects of array batch, dye-swap, and hybridization date we fitted separate spot-wise linear models to the data from our study for each of these three parameters. Gaussian kernel densities were then plotted for the coefficients of the fits for each model. Each curve is unimodal, with a sharp peak around the origin, indicating that most spots on the array do not vary with these design parameters. The spot counts along the bottom of the figure further quantify the degree of variability at several thresholds (0.5= 1.4-fold, 1.0 = 2.0-fold, 1.5 = 2.8-fold, 2. 0= 4.0-fold).

Figure 4: Comparing DNA amplification methods

To evaluate whether specific DNA amplification protocols introduce different degrees of bias, we surveyed the genome-wide effects of three amplification protocols. A) Total

genomic DNA was amplified with either random priming (RP-PCR) or one of two LM-PCR protocols (LM-PCR-A and LM-PCR-B). The resulting DNA was then hybridized to a CpG Island microarray against an equal amount of unamplified total genomic DNA. This procedure was repeated four times for each of the three amplification methods, and the raw data was pre-processed with VSN and model-based t-tests (see Methods). Black arrows indicate reciprocal labeling. B) The fraction of spots on the array that are differentially amplified (fraction of spots called) is plotted against the p-value threshold for each amplification method. The curve for the first LM-PCR protocol (LM-PCR A) lies below that of the other two methods for all p-values, demonstrating that this method has the smallest amplification bias of the three protocols tested.

Figure 5: Comparing hybridization controls

A number of potential control samples can be used for ChIP-chip studies. (A) Three common approaches were compared, including two direct hybridization approaches (No Ab and IgG) and an indirect hybridization approach (No Ab). For each approach ChIP-chip studies of Myc in HL60 cells were performed. Black arrows indicate reciprocal labeling. (B) The sensitivity of each approach was plotted as the percentage of spots identified as differentially bound vs. the p-value. While the three methods displayed similar performance, the two direct methods show higher sensitivity than the indirect method.

Figure 6: Myc recruitment to regulatory regions of novel target genes in growing HL60 cells.

ChIP assays with antibody to Myc and IgG control antibody from exponentially growing HL60 cells. DNA was Q-PCR amplified by using primers against a positive (Cyclin D2) and a negative control (Chromosome 21 E-Box), as well as primers against five sequences from the top third of the list of hits, primers against four sequences from the middle third of the list of hits, and primers against four selected from the bottom third of the list of hits. Array Rank was determined by p-values. *GDAP2*: Ganglioside Induced Differentiation Associated Protein 2. *RPL34*: Ribosomal Protein 34. *PHLPP*: PH Domain and Leucine Rich Repeat Protein Phosphatase. *TCERG1*: Transcription Elongation Regulator 1. *PKIB*: Protein Kinase (cAMP-dependent, catalytic) Inhibitor Beta. *CDC14A*: Cell Division Cycle 14A. OGG1: 8-oxoguanine DNA Glycosylase. *CDCA2*: Cell Division Cycle Associated 2. *NOL6*: Nucleolar Protein Family 6 (RNA-associated).

PLTP: Phospholipid Transfer Protein. *TRH*: Thyrotropin-Releasing Hormone. *KLC4*: Kinesin Light Chain 4. *GAS1*: Growth Arrest-Specific 1.
(1) and **(2)** indicates two primer pairs against different regions of the designated target gene promoter.
The mean log2|Myc/IgG| values are plotted for each target gene, and the Chr21 E-box. Error bars represent standard-deviations. Targets with $p < 0.05$ are marked with a *, statistical significance as assessed using a paired *t* test.

Please be aware that 3 additional data files will be included with the manuscript: Supplementary tables 1, 2 and 3

Figure 1

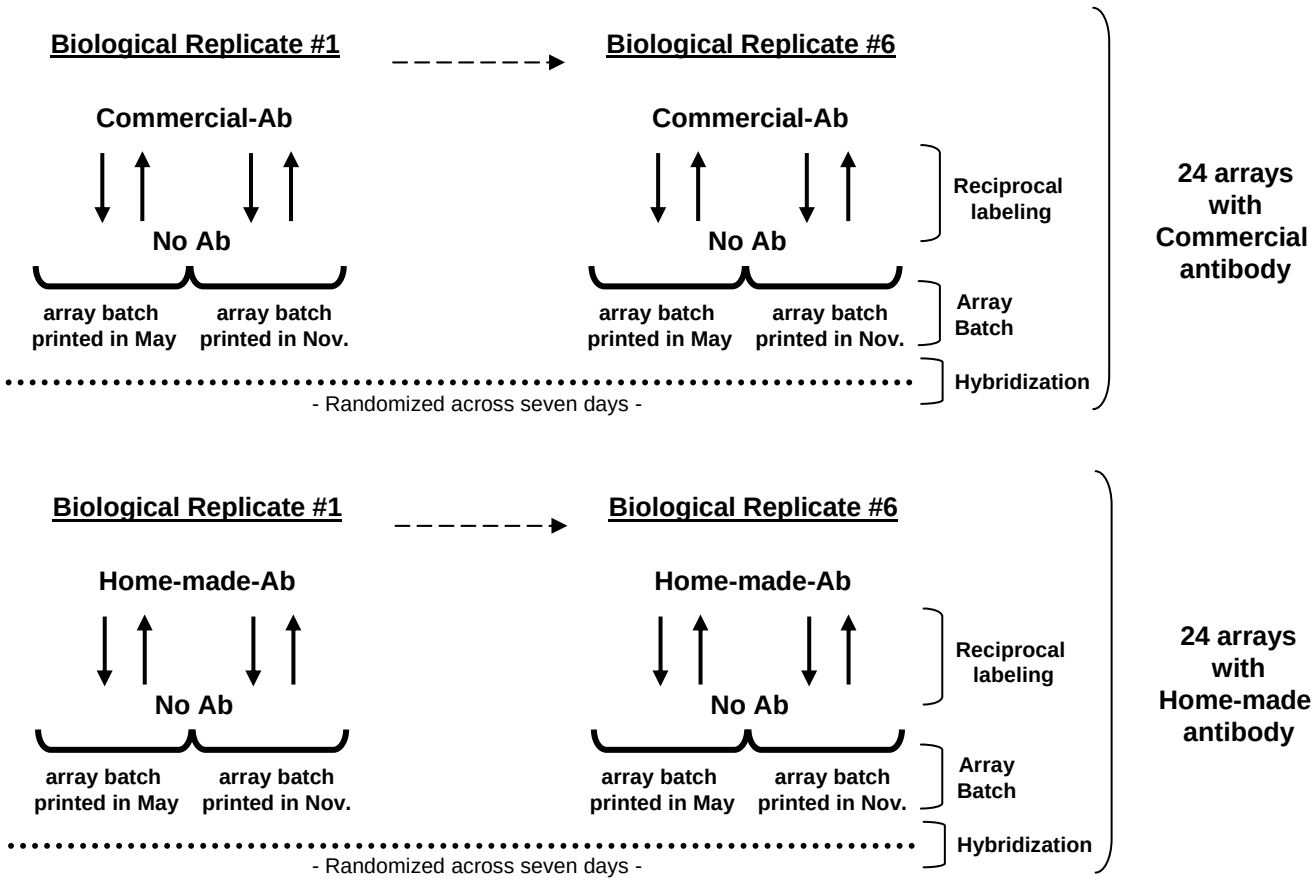

Biological Replicate #1
Biological Replicate #6
Commercial-Ab
Commercial-Ab
No Ab
No Ab
Reciprocal labeling
array batch printed in May
array batch printed in Nov.
array batch printed in May
array batch printed in Nov.
Array Batch
Hybridization
- Randomized across seven days -
24 arrays with Commercial antibody
Biological Replicate #1
Biological Replicate #6
Home-made-Ab
Home-made-Ab
No Ab
No Ab
Reciprocal labeling
array batch printed in May
array batch printed in Nov.
array batch printed in May
array batch printed in Nov.
Array Batch
Hybridization
- Randomized across seven days -
24 arrays with Home-made antibody

Figure 2

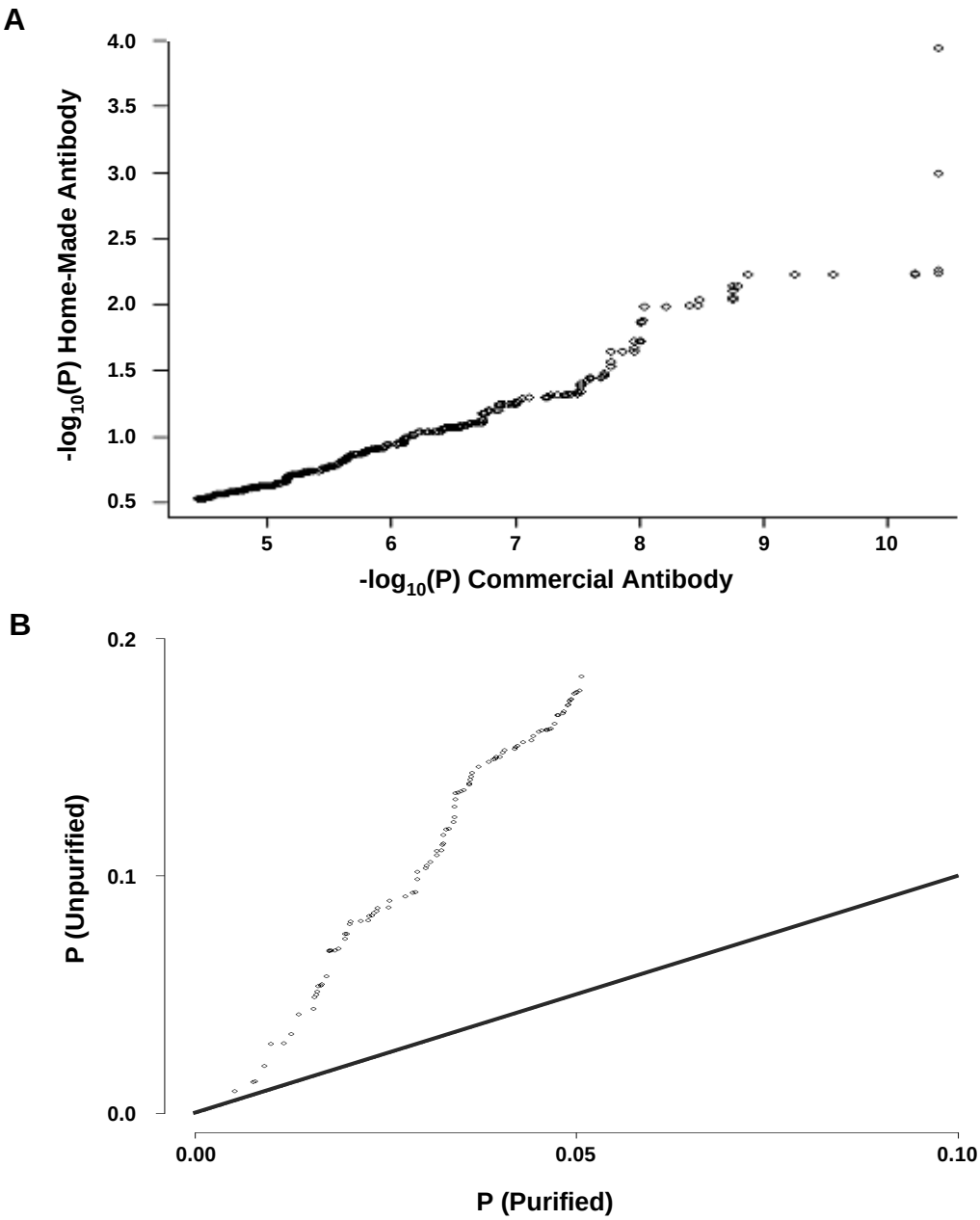
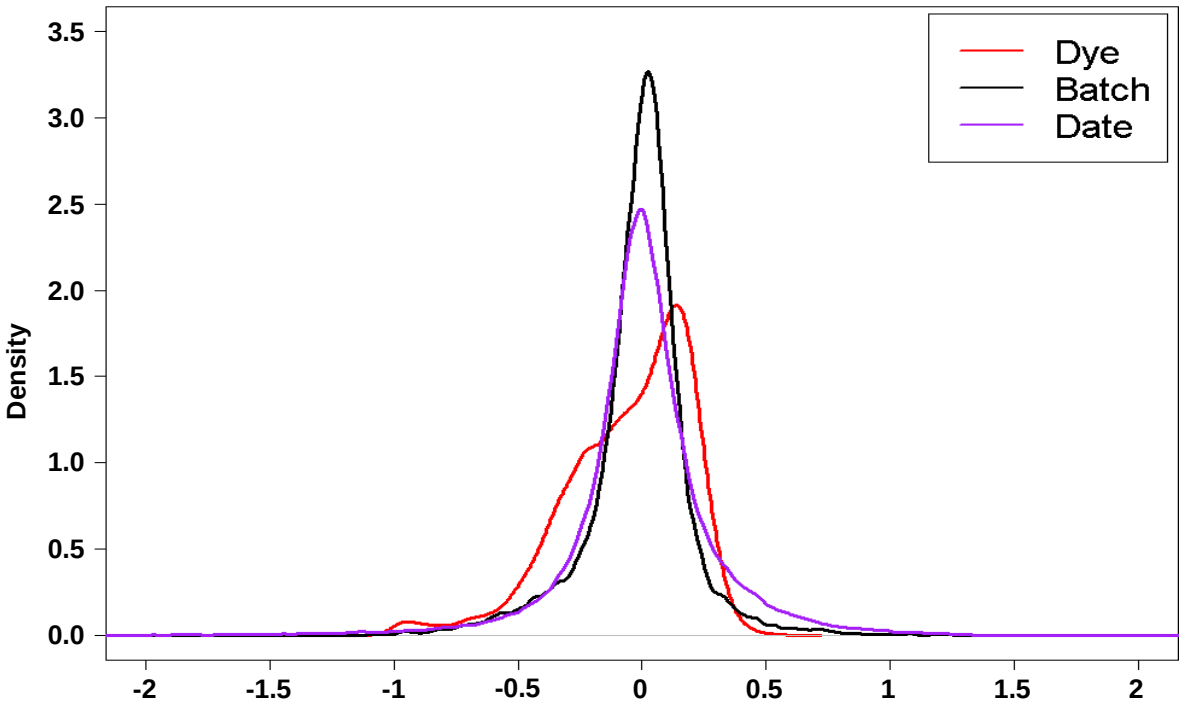

A
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
-log10(P) Home-Made Antibody
5
6
7
8
9
10
-log10(P) Commercial Antibody
B
0.2
0.1
0.0
P (Unpurified)
0.00
0.05
0.10
P (Purified)

Figure 3



Dye	0	0	16	685	-	11	0	0	0
Batch	0	1	13	408	-	210	30	5	0
Date	1	50	147	540	-	575	67	5	1

Spot Counts

Figure 4

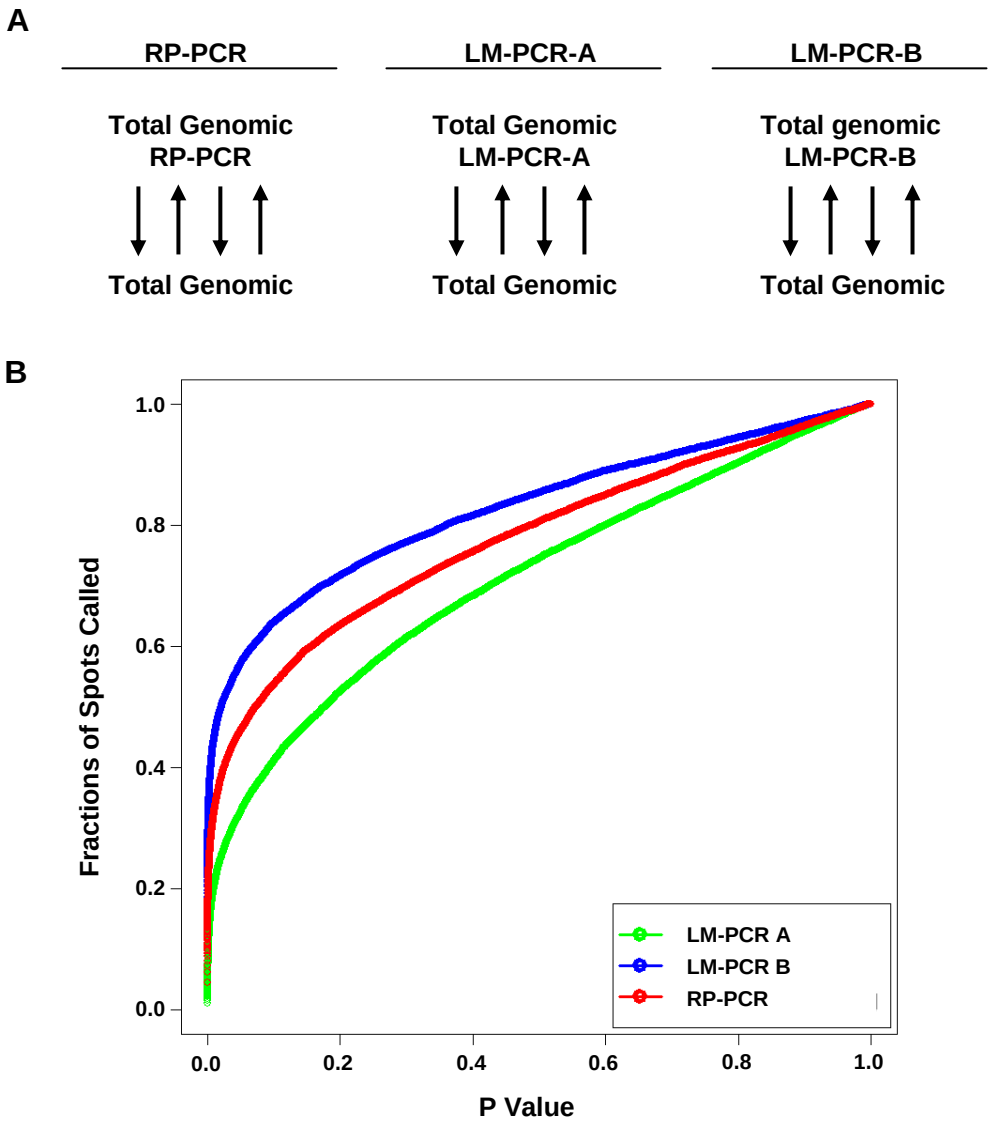

A
RP-PCR
LM-PCR-A
LM-PCR-B
Total Genomic
RP-PCR
Total Genomic
LM-PCR-A
Total genomic
LM-PCR-B
Total Genomic
Total Genomic
Total Genomic
B
Fractions of Spots Called
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
P Value
LM-PCR A
LM-PCR B
RP-PCR

Figure 5

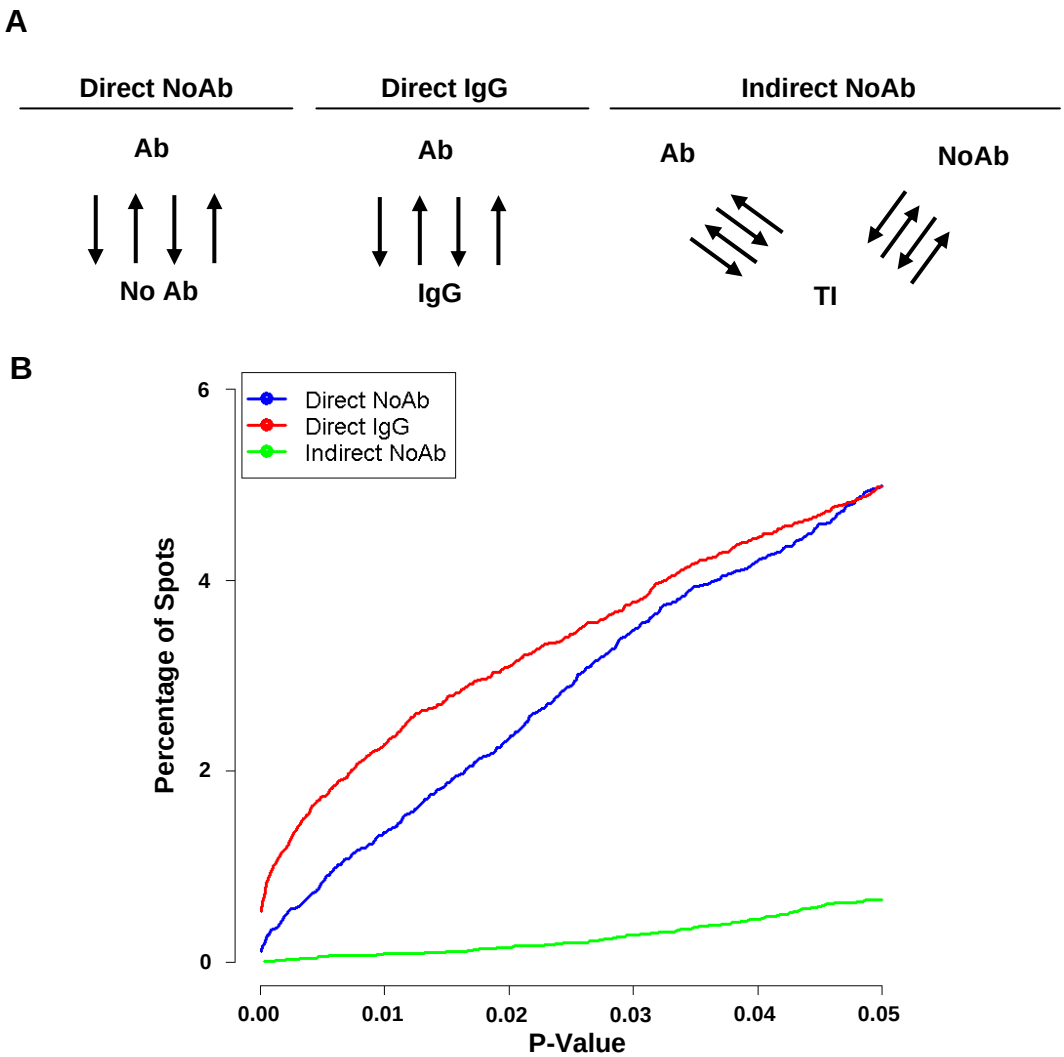
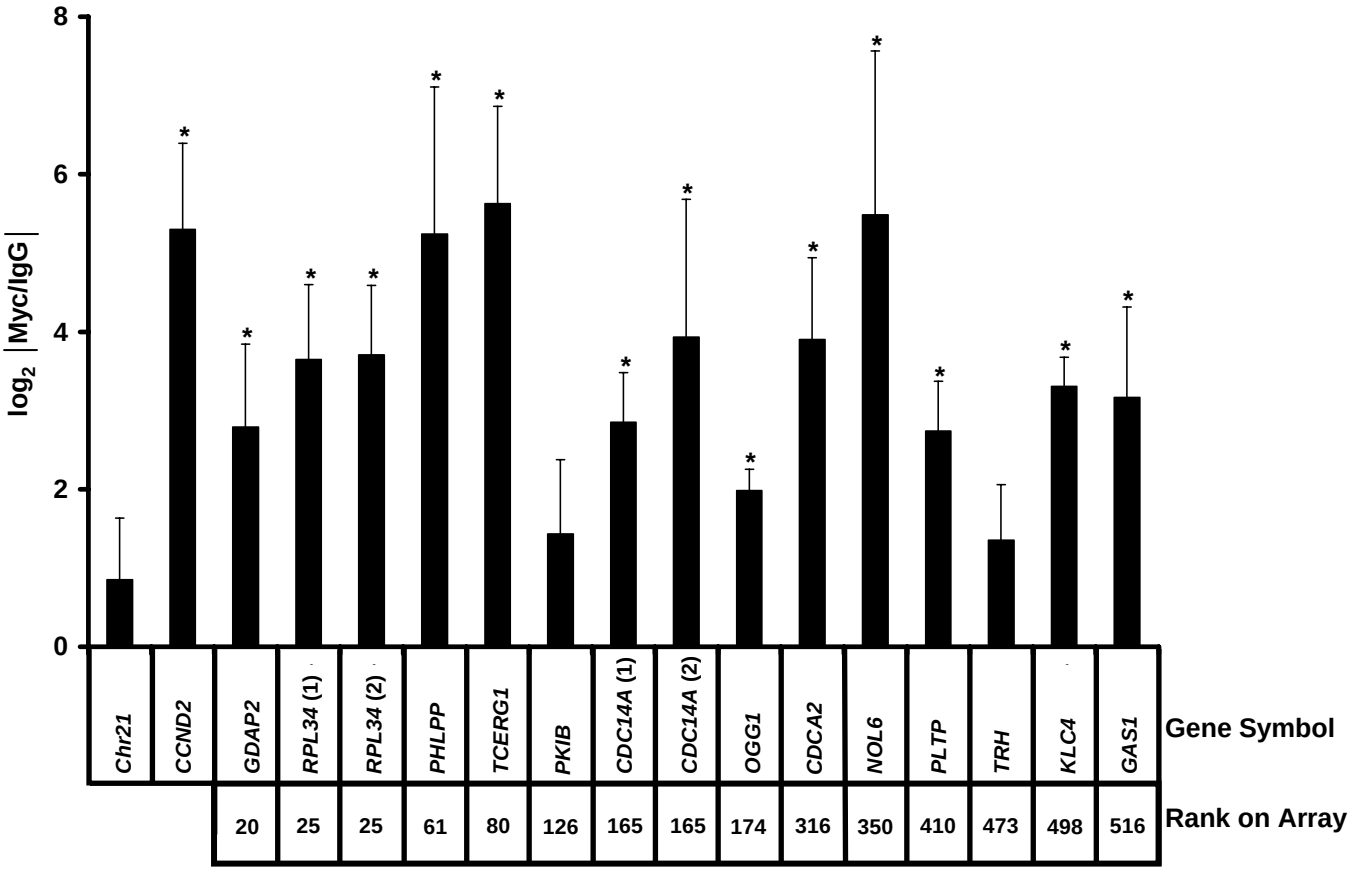

A
Direct NoAb
Direct IgG
Indirect NoAb
Ab
Ab
Ab
NoAb
No Ab
IgG
TI
B
Direct NoAb
Direct IgG
Indirect NoAb
Percentage of Spots
0
2
4
6
0.00
0.01
0.02
0.03
0.04
0.05
P-Value

Figure 6


log2 |Myc/IgG|
8
6
4
2
0
Chr21
CCND2
GDAP2
RPL34 (1)
RPL34 (2)
PHLPP
TCERG1
PKIB
CDC14A (1)
CDC14A (2)
OGG1
CDCA2
NOL6
PLTP
TRH
KLC4
GAS1
Gene Symbol
20
25
25
61
80
126
165
165
174
316
350
410
473
498
516
Rank on Array

Supplementary Table 1: Myc bound spots

Sequence ID	P Values						
	Myc	Specificity Score	Total Hits	Location	Length	Gene1 Distance	Gene1 Direction
UHNhscpg0002222	0.00001	0.845	4	chr19:32976037-32976943	876	1419011	Upstream
UHNhscpg0001035	0.00004	0.257	3	chr7:2741354-2741835	478	55666	Upstream
UHNhscpg0001586	0.00004	0.579	3	chr9:83397628-83398178	503	95238	Upstream
UHNhscpg0006733	0.00004	1	1	chr1:19450425-19451233	761	0	Overlapping
UHNhscpg0001800	0.00005	1	1	chr19:50373163-50373631	457	0	Overlapping
UHNhscpg0000804	0.00006	0.832	2	chr12:56451719-56452193	452	0	Overlapping
UHNhscpg0002414	0.00008	0.84	4	chr19:32976037-32976919	852	1419035	Upstream
UHNhscpg0005461	0.00008	0.196	20	chr12:8270998-8271748	751	0	Overlapping
UHNhscpg0006220	0.00008	1	1	chr19:5741465-5742061	581	129	Upstream
UHNhscpg0000561	0.0001	1	1	chr12:10765755-10766442	663	729	Upstream
UHNhscpg0002303	0.00011	1	1	chr9:2147634-2148121	484	142293	Upstream
UHNhscpg0001085	0.00015	0.908	6	chr19:49308930-49309510	573	0	Overlapping
UHNhscpg0002710	0.00017	1	1	chr5:11022787-11023385	574	208448	Downstream
UHNhscpg0001460	0.0002	1	1	chr19:5742064-5742385	310	0	Overlapping
UHNhscpg0001704	0.00022	0.576	3	chr19:7049822-7050201	368	29352	Upstream
UHNhscpg0001074	0.00024	1	1	chr20:2398526-2399037	500	462	Upstream
UHNhscpg0011585	0.00024	0.424	84	chr19:20995209-20995601	349	0	Overlapping
UHNhscpg0002018	0.00026	0.92	3	chr17:24071016-24071497	451	0	Overlapping
UHNhscpg0005990	0.00028	0.512	2	chr13:40666276-40666857	578	0	Overlapping
UHNhscpg0008598	0.00028	0.063	2	chr1:118273051-118273526	476	203	Upstream
UHNhscpg0000995	0.0003	0.038	2	chr20:24612471-24612553	79	214637	Upstream
UHNhscpg0009461	0.0003	0	8	NA			NA
UHNhscpg0002618	0.00035	0	2	chr9:5293809-5294483	404	97	Upstream
UHNhscpg0001907	0.00036	0.268	13	chr9:44341816-44342357	471	1217010	Upstream
UHNhscpg0002554	0.00038	0.943	2	chr4:109760868-109761756	877	0	Overlapping
UHNhscpg0007878	0.00038	0.892	6	chr19:44618067-44618707	480	0	Overlapping
UHNhscpg0001092	0.00039	1	1	chr2:15649023-15649865	803	0	Overlapping
UHNhscpg0002428	0.00054	1	1	chr5:11022787-11023383	572	208448	Downstream
UHNhscpg0007553	0.00056	0.869	2	chr9:2005281-2005843	518	0	Overlapping
UHNhscpg0003980	0.00061	1	1	chr3:20137200-20137313	114	65374	Upstream
UHNhscpg0000036	0.00071	1	1	chr11:22171599-22172151	532	302	Upstream
UHNhscpg0000958	0.00071	0	43	NA			NA
UHNhscpg0002341	0.00075	0.624	8	chr19:63022721-63023265	521	4628	Downstream
UHNhscpg0000901	0.00079	1	1	chr6:2854867-2854934	64	6361	Downstream
UHNhscpg0008954	0.00079	1	1	chr17:43975297-43975731	435	1661	Upstream
UHNhscpg0009653	0.00079	0.711	2	chr6:96131747-96132044	294	89	Downstream
UHNhscpg0001188	0.0008	0.006	12	chr9:39807205-39808037	695	66937	Downstream
UHNhscpg0001063	0.00083	1	1	chr15:73533098-73533556	459	2119	Downstream
UHNhscpg0001735	0.00089	0.926	2	chr19:750081-750911	811	1671	Upstream
UHNhscpg0002090	0.00091	1	1	chr16:1300396-1300886	444	758	Upstream
UHNhscpg0002327	0.00098	1	1	chr6:136652761-136653188	388	79	Downstream
UHNhscpg0001703	0.00099	1	1	chr6:30289803-30290357	539	671	Downstream
UHNhscpg0001014	0.00103	1	1	chr11:62115235-62115799	561	0	Overlapping
UHNhscpg0001248	0.00106	1	1	chr10:97880130-97880692	555	669	Upstream
UHNhscpg0002390	0.00108	0.713	2	chr16:20724860-20725654	774	0	Overlapping
UHNhscpg0001860	0.00116	1	1	chr1:210274778-210276110	1258	0	Overlapping
UHNhscpg0002613	0.00119	0.882	4	chr5:138920411-138920816	382	118	Downstream
UHNhscpg0002892	0.00123	0.921	2	chr5:180603052-180603713	647	0	Overlapping
UHNhscpg0001696	0.00126	0.382	2	chr7:16651379-16651586	204	337	Upstream
UHNhscpg0010670	0.00127	0	9	NA			NA
UHNhscpg0004835	0.00129	0.084	3	chr17:75426047-75426328	226	1480	Upstream
UHNhscpg0002522	0.0013	1	1	chr5:11022787-11023385	574	208448	Downstream
UHNhscpg0001893	0.00133	0.928	2	chr21:39674027-39674584	554	0	Overlapping
UHNhscpg0006792	0.00136	1	1	chr12:124039302-124039784	458	0	Overlapping
UHNhscpg0000307	0.00146	0.792	2	chr6:10803142-10803420	279	0	Overlapping

Sequence ID	P Values	Specificity Score	Total Hits	Location	Length	Gene1 Distance	Gene1 Direction
	Myc						
UHNhscpg0002423	0.00148	0.514	122	chr19:21333537-21334178	622	29485	Downstream
UHNhscpg0010053	0.00151	0	63	NA			NA
UHNhscpg0008201	0.0016	0.395	3	chr2:122004252-122004648	397	118874	Upstream
UHNhscpg0000807	0.00162	1	1	chr6:2854864-2854934	71	6358	Downstream
UHNhscpg0006450	0.00162	0.941	2	chr5:37406425-37407113	682	0	Overlapping
UHNhscpg0004878	0.00166	1	1	chr18:58533165-58533572	359	141	Downstream
UHNhscpg0005565	0.0017	0.119	3	chr12:31117908-31118439	528	0	Overlapping
UHNhscpg0009740	0.00171	0.085	2	chr1:27564765-27564974	142	950	Upstream
UHNhscpg0001762	0.00172	1	1	chr4:140155954-140156234	269	158	Downstream
UHNhscpg0007633	0.00173	1	1	chr1:21982660-21982960	298	385	Downstream
UHNhscpg0000400	0.00174	1	1	chr6:90586005-90586465	461	0	Overlapping
UHNhscpg0002435	0.00181	0.903	2	chr15:61236276-61236994	627	0	Overlapping
UHNhscpg0006438	0.00181	0.097	30	chr7:64872716-64873578	826	0	Overlapping
UHNhscpg0011811	0.00181	1	1	chr2:201099159-201099469	303	0	Overlapping
UHNhscpg0010108	0.00183	0.16	2	chr1:31998811-31999148	338	3087	Upstream
UHNhscpg0002065	0.00184	1	1	chr17:69711629-69712261	525	240	Upstream
UHNhscpg0001877	0.00187	0	67	NA			NA
UHNhscpg0006539	0.0019	1	1	chr5:79322136-79322956	821	0	Overlapping
UHNhscpg0000856	0.00199	0.799	4	chr19:49220922-49221654	725	0	Overlapping
UHNhscpg0004517	0.00202	0.705	8	chr13:40393343-40393901	559	60517	Upstream
UHNhscpg0004234	0.00203	1	1	chr7:104439994-104441046	1049	826	Downstream
UHNhscpg0001340	0.00209	1	1	chr3:72232401-72232969	545	315499	Downstream
UHNhscpg0001999	0.00214	0.181	13	chr9:44341816-44342348	415	1217010	Upstream
UHNhscpg0001718	0.00221	0.102	4	chr6:26853557-26853991	431	85816	Downstream
UHNhscpg0003837	0.00222	1	1	chr5:145806606-145807299	686	0	Overlapping
UHNhscpg0002637	0.00223	0.523	2	chr12:48421912-48422646	707	9865	Downstream
UHNhscpg0011517	0.00223	1	1	chr19:11061002-11061429	407	0	Overlapping
UHNhscpg0002077	0.00232	1	1	chr6:90586005-90586469	461	0	Overlapping
UHNhscpg0002216	0.00236	1	1	chr12:52959902-52960341	413	0	Overlapping
UHNhscpg0001195	0.00237	0.024	2	chr19:50620370-50620427	42	1353	Downstream
UHNhscpg0007470	0.00237	0.527	2	chr1:32168088-32168241	74	8334	Upstream
UHNhscpg0002483	0.00243	0.296	3	chr16:85145935-85146230	280	112	Upstream
UHNhscpg0001574	0.00256	1	1	chr6:144458002-144458718	713	0	Overlapping
UHNhscpg0001709	0.00257	0.922	2	chr3:44664986-44665759	770	0	Overlapping
UHNhscpg0006569	0.00259	1	1	chr11:111254731-111255533	803	0	Overlapping
UHNhscpg0001158	0.00265	1	1	chr17:39935622-39936669	1024	0	Overlapping
UHNhscpg0002640	0.0027	1	1	chr2:85501234-85502047	810	10047	Downstream
UHNhscpg0000920	0.00282	1	1	chr11:62115235-62115803	565	0	Overlapping
UHNhscpg0006739	0.00282	0.236	2	chr1:3653933-3654076	144	883	Downstream
UHNhscpg0001814	0.00288	0.268	13	chr9:44341816-44342357	471	1217010	Upstream
UHNhscpg0002399	0.0029	1	1	chr5:140730098-140730485	384	0	Overlapping
UHNhscpg0002080	0.00299	0.187	2	chr19:13746170-13746588	395	0	Overlapping
UHNhscpg0001890	0.00302	0.869	2	chr12:51975674-51976234	534	173	Upstream
UHNhscpg0006364	0.00302	1	1	chr19:9740462-9740746	285	52	Downstream
UHNhscpg0003776	0.00305	1	1	chrX:38545425-38545774	350	2242	Downstream
UHNhscpg0000897	0.00317	0	66	NA			NA
UHNhscpg0009677	0.00318	0.871	2	chr17:72065094-72065566	457	17328	Upstream
UHNhscpg0006124	0.00332	0.271	2	chr9:95748240-95748309	70	9114	Upstream
UHNhscpg0000478	0.00335	0.797	2	chr2:113015948-113016337	390	0	Overlapping
UHNhscpg0001754	0.00347	0.897	2	chr17:78155108-78155715	400	43851	Upstream
UHNhscpg0002457	0.0035	0.752	3	chr5:125958442-125959093	636	0	Overlapping
UHNhscpg0008746	0.00353	1	1	chr1:164064064-164064435	225	551	Upstream
UHNhscpg0001954	0.0036	0	12	NA			NA
UHNhscpg0002604	0.0036	0.85	3	chr19:32976290-32976943	439	1419011	Upstream
UHNhscpg0004396	0.00369	0.935	2	chr18:55090106-55091038	922	567	Upstream
UHNhscpg0008244	0.00375	0.897	2	chr4:84175025-84175531	507	0	Overlapping
UHNhscpg0001908	0.00377	1	1	chr12:119239392-119240071	665	9904	Upstream

Sequence ID	P Values	Specificity Score	Total Hits	Location	Length	Gene1 Distance	Gene1 Direction
	Myc						
UHNhscpg0002101	0.0038	0.694	2	chr9:35106261-35106931	647	368	Downstream
UHNhscpg0002273	0.0038	1	1	chr12:102489049-102489597	537	15601	Downstream
UHNhscpg0011595	0.00385	0	35	NA			NA
UHNhscpg0001182	0.0039	0	13	NA			NA
UHNhscpg0009185	0.0039	0	5	chr1:148070730-148071003	266	0	Overlapping
UHNhscpg0005267	0.00406	1	1	chr19:12695458-12696086	629	1380	Downstream
UHNhscpg0007192	0.00413	1	1	chr6:24829288-24829774	463	1906	Downstream
UHNhscpg0010640	0.00414	1	1	chr16:81828084-81828454	367	570639	Downstream
UHNhscpg0000808	0.00424	1	1	chr7:45117391-45118238	844	0	Overlapping
UHNhscpg0001315	0.00435	1	1	chr2:207338359-207338723	353	64	Downstream
UHNhscpg0006036	0.00435	1	1	chr22:40347412-40347900	489	172	Upstream
UHNhscpg0006692	0.0044	0	39	NA			NA
UHNhscpg0000663	0.00441	0	4	chr7:143029357-143029674	94	691	Upstream
UHNhscpg0008592	0.00454	1	1	chr6:122972820-122973031	132	44	Downstream
UHNhscpg0002718	0.00459	1	1	chr4:53926306-53927018	701	0	Overlapping
UHNhscpg0001461	0.00466	1	1	chr11:124241472-124242010	527	981	Upstream
UHNhscpg0011747	0.00466	1	1	chr13:46268564-46268875	308	493	Upstream
UHNhscpg0004015	0.00489	1	1	chr2:107581563-107581793	219	387633	Downstream
UHNhscpg0002158	0.00491	0.762	2	chr2:127999967-128000281	311	233	Upstream
UHNhscpg0004300	0.00491	1	1	chr13:35769917-35770383	467	48070	Upstream
UHNhscpg0004008	0.00495	1	1	chr12:119503184-119504013	810	0	Overlapping
UHNhscpg0008423	0.00501	0.883	2	chr6:116998749-116999278	530	0	Overlapping
UHNhscpg0008176	0.00502	0.458	3	chr3:47299103-47299615	513	0	Overlapping
UHNhscpg0009268	0.00504	0.909	2	chr2:73817431-73818211	769	0	Overlapping
UHNhscpg0004645	0.00508	1	1	chr6:43089334-43090121	758	0	Overlapping
UHNhscpg0002437	0.00514	0.624	8	chr19:63022721-63023265	521	4628	Downstream
UHNhscpg0000389	0.00519	1	1	chr9:94095629-94096224	572	0	Overlapping
UHNhscpg0005425	0.00523	1	1	chr22:36533363-36533799	429	101	Downstream
UHNhscpg0002339	0.00526	0.908	2	chr15:61236270-61236994	665	0	Overlapping
UHNhscpg0004394	0.00533	0.357	4	chr19:57334937-57335121	185	0	Overlapping
UHNhscpg0003606	0.00537	1	1	chr8:96350447-96351060	614	0	Overlapping
UHNhscpg0009520	0.0054	1	1	chr1:91738839-91739081	243	0	Overlapping
UHNhscpg0008070	0.00549	1	1	chr1:191490112-191490214	99	67765	Downstream
UHNhscpg0001041	0.00566	1	1	chr1:154808695-154809177	479	0	Overlapping
UHNhscpg0009274	0.0057	0	9	NA			NA
UHNhscpg0005239	0.00582	1	1	chr19:63590114-63590420	264	27	Downstream
UHNhscpg0005459	0.00584	1	1	chr6:153345899-153346717	784	32	Downstream
UHNhscpg0004074	0.00593	0	2	NA			NA
UHNhscpg0006096	0.00594	0.621	3	chr12:94860144-94860524	282	35	Upstream
UHNhscpg0005614	0.00599	0.549	2	chr10:32256976-32257108	133	668	Upstream
UHNhscpg0001279	0.00607	1	1	chr2:15649014-15649865	808	0	Overlapping
UHNhscpg0002006	0.00611	1	1	chr1:68288598-68289125	452	0	Overlapping
UHNhscpg0001133	0.00616	1	1	chr1:154808695-154809177	479	0	Overlapping
UHNhscpg0008635	0.00617	0.339	2	chr2:45008927-45008985	59	13555	Downstream
UHNhscpg0002795	0.0064	1	1	chr19:50601483-50602103	349	177	Upstream
UHNhscpg0010736	0.0064	1	1	chr16:81828084-81828447	360	570646	Downstream
UHNhscpg0000314	0.00645	0.655	5	chr11:85633827-85634078	252	365	Upstream
UHNhscpg0001205	0.00655	0.507	2	chr17:38793349-38793666	302	38212	Downstream
UHNhscpg0001143	0.00659	0	43	NA			NA
UHNhscpg0001237	0.00664	1	1	chr15:67673499-67674364	842	141287	Upstream
UHNhscpg0000480	0.00665	1	1	chr7:135312099-135312915	727	0	Overlapping
UHNhscpg0004701	0.00682	0.757	2	chr4:3046987-3047167	181	782	Upstream
UHNhscpg0001967	0.00689	1	1	chr1:100589367-100589715	345	919	Downstream
UHNhscpg0000915	0.00691	0.106	2	chr19:50619268-50619314	47	251	Downstream
UHNhscpg0005276	0.00704	0.74	2	chr6:30646907-30647379	473	0	Overlapping
UHNhscpg0001848	0.00705	0	34	NA			NA
UHNhscpg0001795	0.00714	1	1	chr22:45536661-45537115	431	97	Downstream

Sequence ID	P Values	Specificity Score	Total Hits	Location	Length	Gene1 Distance	Gene1 Direction
	Myc						
UHNhscpg0008428	0.00717	0	2	NA			NA
UHNhscpg0004303	0.00718	0.563	33	chr6:27883373-27883888	508	0	Overlapping
UHNhscpg0001809	0.00728	0	15	NA			NA
UHNhscpg0000119	0.00737	0.41	13	chr17:25951567-25951939	373	40858	Upstream
UHNhscpg0009385	0.00744	0.572	2	chr3:9765933-9766477	545	176	Downstream
UHNhscpg0000272	0.00749	0.457	4	chr5:131853530-131853938	405	388	Upstream
UHNhscpg0000705	0.00772	0.486	2	chr14:51396904-51397151	210	16908	Downstream
UHNhscpg0000925	0.00779	1	1	chr20:11257476-11257777	235	561699	Downstream
UHNhscpg0000449	0.00785	0	13	NA			NA
UHNhscpg0002569	0.00786	0.761	4	chr19:49337662-49337941	268	45	Upstream
UHNhscpg0000556	0.00787	0.834	2	chr5:89805381-89806166	763	175	Upstream
UHNhscpg0001206	0.0079	0	29	chr1:16865365-16865789	377	52796	Downstream
UHNhscpg0008723	0.00807	0	13	NA			NA
UHNhscpg0008688	0.00814	0.62	3	chr1:237616558-237616870	305	242125	Downstream
UHNhscpg0001530	0.00823	1	1	chr7:13995333-13995759	251	44	Downstream
UHNhscpg0003878	0.00823	1	1	chr2:238539828-238540325	474	113	Downstream
UHNhscpg0001947	0.00824	0.784	2	chr16:14073060-14073523	444	364	Upstream
UHNhscpg0009557	0.00831	0.711	2	chr6:96131747-96132044	294	89	Downstream
UHNhscpg0009765	0.00834	0	63	NA			NA
UHNhscpg0001955	0.00848	1	1	chr1:210274778-210276110	1234	0	Overlapping
UHNhscpg0002254	0.00853	0.681	2	chr9:7789295-7789855	524	0	Overlapping
UHNhscpg0011787	0.00857	1	1	chr6:106548952-106549065	114	91822	Downstream
UHNhscpg0000048	0.00864	0.614	3	chr18:31177840-31178483	621	0	Overlapping
UHNhscpg0002200	0.00864	1	1	chr14:76993362-76994131	758	0	Overlapping
UHNhscpg0000491	0.00873	0.965	2	chr6:159340256-159341432	1121	0	Overlapping
UHNhscpg0002926	0.00886	0	43	NA			NA
UHNhscpg0000927	0.00887	1	1	chr17:70157894-70158326	425	20538	Downstream
UHNhscpg0001162	0.0091	0	23	NA			NA
UHNhscpg0005505	0.00912	1	1	chr16:31134123-31135151	994	745	Upstream
UHNhscpg0005734	0.00917	0	35	NA			NA
UHNhscpg0000278	0.00923	1	1	chr3:48647313-48648039	727	0	Overlapping
UHNhscpg0002243	0.00929	0.912	2	chr15:61236270-61236994	697	0	Overlapping
UHNhscpg0002446	0.0093	0	13	NA			NA
UHNhscpg0011406	0.0093	1	1	chr11:59075109-59075158	50	36148	Upstream
UHNhscpg0010874	0.00933	1	1	chr13:46268564-46268875	308	493	Upstream
UHNhscpg0000286	0.00943	1	1	chr6:30748320-30748935	529	0	Overlapping
UHNhscpg0007078	0.00949	1	1	chr5:115317061-115317139	79	9030	Downstream
UHNhscpg0006940	0.0095	1	1	chr19:9295687-9295872	170	55	Downstream
UHNhscpg0011843	0.00966	1	1	chr13:46268564-46268875	308	493	Upstream
UHNhscpg0001113	0.00993	0.476	2	chr7:20783714-20784421	656	8609	Upstream
UHNhscpg0001527	0.00997	1	1	chr20:8996180-8996731	541	28200	Downstream
UHNhscpg0000168	0.01009	0.085	2	chr13:99431681-99431833	153	486	Downstream
UHNhscpg0001933	0.01012	0.263	2	chr15:80342224-80342711	380	18	Upstream
UHNhscpg0000809	0.01015	0.823	5	chr19:62614206-62614920	711	0	Overlapping
UHNhscpg0000882	0.01027	0	17	NA			NA
UHNhscpg0009269	0.01029	1	1	chr2:218940536-218941148	597	10714	Upstream
UHNhscpg0001322	0.01035	0	2	chr5:69381228-69381607	376	123	Upstream
UHNhscpg0000372	0.01038	1	1	chr4:76868579-76869553	939	0	Overlapping
UHNhscpg0009193	0.01047	1	1	chr10:89567438-89567728	239	169	Upstream
UHNhscpg0001154	0.01053	1	1	chr17:17080205-17080795	587	426	Upstream
UHNhscpg0001888	0.01055	1	1	chr22:45536661-45537115	431	97	Downstream
UHNhscpg0001569	0.01057	1	1	chr6:138230630-138231068	427	357	Upstream
UHNhscpg0001578	0.01076	1	1	chr18:73075889-73076383	483	14337	Downstream
UHNhscpg0001506	0.01078	0.607	2	chr13:26723246-26723668	415	23	Downstream
UHNhscpg0000937	0.01094	0	19	NA			NA
UHNhscpg0007203	0.011	0.746	2	chr1:35317368-35317832	456	0	Overlapping
UHNhscpg0000658	0.01118	0.865	2	chr10:79458886-79459376	436	0	Overlapping

Sequence ID	P Values	Specificity Score	Total Hits	Location	Length	Gene1 Distance	Gene1 Direction
	Myc						
UHNhscpg0003155	0.01122	0.033	35	chr4:49209349-49209607	215	526328	Upstream
UHNhscpg0006329	0.01128	0.453	9	chr1:67668270-67668930	653	0	Overlapping
UHNhscpg0008734	0.01146	1	1	chr10:122729028-122729311	280	128344	Upstream
UHNhscpg0007212	0.01183	0.858	2	chr17:34983441-34984081	641	33620	Upstream
UHNhscpg0002673	0.01195	1	1	chr15:81825125-81825456	235	80653	Downstream
UHNhscpg0001300	0.012	0	2	chr5:69381228-69381607	372	123	Upstream
UHNhscpg0001291	0.01203	1	1	chr11:62115235-62115799	479	0	Overlapping
UHNhscpg0000596	0.0122	0.872	3	chr11:960138-961029	797	44298	Upstream
UHNhscpg0011233	0.0123	0	47	NA			NA
UHNhscpg0001228	0.01239	1	1	chr1:154808695-154809177	479	0	Overlapping
UHNhscpg0011824	0.01243	0.847	2	chr16:2144955-2145183	229	616	Downstream
UHNhscpg0005883	0.01259	1	1	chr1:10381074-10382309	1229	0	Overlapping
UHNhscpg0005465	0.01273	1	1	chr6:71049050-71049719	670	0	Overlapping
UHNhscpg0000441	0.01291	0.567	2	chr13:40604530-40605211	670	0	Overlapping
UHNhscpg0002181	0.01293	0.494	2	chr8:132985502-132985928	427	35401	Downstream
UHNhscpg0004594	0.01294	1	1	chr1:160734111-160734402	255	0	Overlapping
UHNhscpg0002608	0.01299	0.925	3	chr9:116389649-116390322	666	0	Overlapping
UHNhscpg0001075	0.01308	1	1	chr6:27594840-27595208	357	45982	Downstream
UHNhscpg0007619	0.01313	1	1	chr16:84144824-84145260	433	59164	Downstream
UHNhscpg0002828	0.01322	0.183	2	chr22:25086487-25086628	142	68673	Downstream
UHNhscpg0004797	0.01325	1	1	chr4:85722413-85722792	380	288	Downstream
UHNhscpg0009903	0.01325	0.851	3	chr11:77026049-77026697	649	0	Overlapping
UHNhscpg0001122	0.01327	0.231	2	chr4:153186392-153186612	221	284796	Downstream
UHNhscpg0001657	0.01331	1	1	chr3:182113380-182113899	504	235	Upstream
UHNhscpg0001302	0.01335	0.199	4	chr12:47531849-47532218	362	6	Upstream
UHNhscpg0001222	0.01337	0.485	6	chr1:147423176-147423307	132	18539	Downstream
UHNhscpg0001628	0.01339	0.021	10	chr6:58249504-58249834	331	145821	Upstream
UHNhscpg0001391	0.01345	0	29	chr1:16865365-16865789	377	52796	Downstream
UHNhscpg0001120	0.01349	0.526	2	chr3:116859797-116860278	458	34656	Upstream
UHNhscpg0000173	0.01379	1	1	chr4:100703869-100704376	504	21	Downstream
UHNhscpg0007087	0.01409	1	1	chr19:51983258-51983854	597	0	Overlapping
UHNhscpg0000825	0.0143	1	1	chr11:62115235-62115799	561	0	Overlapping
UHNhscpg0007223	0.0147	0.476	3	chr11:35117168-35117692	521	176	Upstream
UHNhscpg0010071	0.01472	1	1	chr10:91960146-91960343	173	508800	Upstream
UHNhscpg0003919	0.01477	1	1	chr2:107581552-107581794	243	387632	Downstream
UHNhscpg0011489	0.01482	0.414	83	chr19:20995209-20995601	389	0	Overlapping
UHNhscpg0002088	0.01512	0.131	11	chr2:71210238-71210899	650	0	Overlapping
UHNhscpg0003021	0.0156	1	1	chr5:56242005-56242320	196	1146	Upstream
UHNhscpg0001776	0.01569	0.607	2	chr13:26723246-26723668	415	23	Downstream
UHNhscpg0004258	0.01584	1	1	chr8:128815427-128815697	271	1800	Downstream
UHNhscpg0002269	0.01591	1	1	chr16:18719891-18721065	1151	0	Overlapping
UHNhscpg0008332	0.01603	0	2	NA			NA
UHNhscpg0002738	0.01607	0.183	2	chr22:25086487-25086628	142	68673	Downstream
UHNhscpg0005886	0.01634	0	106	NA			NA
UHNhscpg0006144	0.01637	1	1	chr12:55148527-55148849	323	0	Overlapping
UHNhscpg0000953	0.01638	0.193	11	chr2:132147266-132148290	1001	47109	Upstream
UHNhscpg0006819	0.01664	1	1	chr1:81711875-81711915	41	326754	Downstream
UHNhscpg0003513	0.01669	1	1	chr8:96350435-96351060	626	0	Overlapping
UHNhscpg0004688	0.01682	1	1	chr8:107738736-107738968	233	301	Downstream
UHNhscpg0001419	0.01688	1	1	chr19:53990095-53990948	819	15165	Upstream
UHNhscpg0004354	0.01698	1	1	chr8:128815427-128815694	256	1803	Downstream
UHNhscpg0001974	0.01712	0.907	11	chr19:60589086-60589859	755	0	Overlapping
UHNhscpg0001103	0.01726	1	1	chr19:11510292-11510795	409	0	Overlapping
UHNhscpg0004192	0.01728	1	1	chr3:25799228-25799413	186	513	Upstream
UHNhscpg0000332	0.01766	0.007	3	chr3:197850403-197850877	446	175	Downstream
UHNhscpg0001298	0.01777	0	29	chr1:16865365-16865789	377	52796	Downstream
UHNhscpg0002046	0.01788	1	1	chr1:210274778-210276110	1289	0	Overlapping

Sequence ID	P Values						
	Myc	Specificity Score	Total Hits	Location	Length	Gene1 Distance	Gene1 Direction
UHNhscpg0000402	0.01803	0.807	3	chr10:70330982-70331591	602	0	Overlapping
UHNhscpg0006392	0.01804	1	1	chr5:10302703-10303007	301	0	Overlapping
UHNhscpg0006618	0.01824	1	1	chr9:4974346-4974895	550	349	Downstream
UHNhscpg0003717	0.01829	1	1	chr2:187058898-187059935	950	0	Overlapping
UHNhscpg0001256	0.01845	0	23	NA			NA
UHNhscpg0002247	0.01848	0.915	2	chr16:30793511-30794178	644	18722	Upstream
UHNhscpg0001008	0.01882	0	11	NA			NA
UHNhscpg0003872	0.01901	1	1	chrX:38545421-38545774	350	2242	Downstream
UHNhscpg0008192	0.01905	1	1	chr10:103532635-103533215	581	0	Overlapping
UHNhscpg0000531	0.01946	1	1	chr4:186171924-186173147	1181	3929	Downstream
UHNhscpg0000072	0.01947	0	23	NA			NA
UHNhscpg0000468	0.01948	0.007	23	chr1:145999017-145999561	541	66522	Upstream
UHNhscpg0000813	0.01957	0.595	3	chr16:28743136-28743416	274	1222	Upstream
UHNhscpg0001492	0.0196	0	12	NA			NA
UHNhscpg0000220	0.01962	0.728	2	chr2:188864982-188865719	738	0	Overlapping
UHNhscpg0002563	0.01971	0.058	5	chr17:18907815-18907957	139	16754	Downstream
UHNhscpg0010709	0.01981	1	1	chr5:31567438-31567801	364	124	Upstream
UHNhscpg0005462	0.02007	0.947	2	chr14:104706380-104707144	749	174	Downstream
UHNhscpg0002145	0.02014	0.461	2	chr1:116762694-116763009	308	45336	Upstream
UHNhscpg0007737	0.02019	1	1	chr11:111254752-111255541	653	0	Overlapping
UHNhscpg0001058	0.02026	1	1	chr19:42021242-42021797	544	121	Downstream
UHNhscpg0010828	0.02046	1	1	chr2:237377856-237377932	55	234738	Upstream
UHNhscpg0002025	0.0206	1	1	chr1:47683502-47683914	393	9227	Upstream
UHNhscpg0000308	0.02079	0.829	2	chr6:121697252-121697969	718	0	Overlapping
UHNhscpg0005353	0.02088	1	1	chr13:72253948-72254560	549	0	Overlapping
UHNhscpg0001906	0.02096	0.749	119	chr15:60324013-60324682	650	3604	Upstream
UHNhscpg0002545	0.02111	0	11	NA			NA
UHNhscpg0009036	0.02111	1	1	chr8:8628184-8628292	95	31109	Upstream
UHNhscpg0001636	0.02115	0.318	2	chr9:17032661-17032855	195	92182	Downstream
UHNhscpg0001536	0.0213	0	43	NA			NA
UHNhscpg0001653	0.0215	0.493	88	chr19:23661339-23661914	537	0	Overlapping
UHNhscpg0008841	0.0215	0.816	2	chr19:60720145-60720879	724	965	Downstream
UHNhscpg0007225	0.02167	1	1	chr8:25372386-25373182	797	0	Overlapping
UHNhscpg0002003	0.02185	0.776	2	chr8:98724881-98725277	393	305	Downstream
UHNhscpg0008297	0.02192	1	1	chr14:89919066-89920156	1069	12969	Downstream
UHNhscpg0009796	0.02195	1	1	chr4:184000942-184001180	239	74444	Upstream
UHNhscpg0004468	0.02197	0.005	4	chr17:41794271-41795059	773	0	Overlapping
UHNhscpg0000824	0.02201	0.033	2	chr11:65412243-65412610	273	0	Overlapping
UHNhscpg0002458	0.02224	0.522	3	chr12:111860306-111860954	602	0	Overlapping
UHNhscpg0004793	0.02228	1	1	chr3:138969216-138969571	356	2948	Upstream
UHNhscpg0002453	0.02271	1	1	chr2:137238469-137238710	122	646274	Downstream
UHNhscpg0001067	0.02292	0	23	NA			NA
UHNhscpg0006728	0.023	1	1	chr6:30247427-30248400	970	7987	Upstream
UHNhscpg0006752	0.02303	1	1	chr14:38971811-38972105	270	440	Downstream
UHNhscpg0006451	0.02308	0.532	61	chr7:6936836-6937722	862	104450	Downstream
UHNhscpg0007253	0.02309	0.004	327	chr1:144034037-144035014	955	89620	Downstream
UHNhscpg0002847	0.02319	0.94	2	chr8:22475190-22475966	765	3193	Downstream
UHNhscpg0000526	0.02346	0.531	27	chr1:147490240-147490606	303	64371	Downstream
UHNhscpg0002450	0.02366	0	11	NA			NA
UHNhscpg0002612	0.02371	0	8	chr1:13332617-13332708	92	35132	Downstream
UHNhscpg0002381	0.02405	0.045	23	chr4:9073118-9073165	44	17761	Upstream
UHNhscpg0009176	0.02406	0.751	10	chr19:63137964-63138541	570	11	Upstream
UHNhscpg0002010	0.02434	0.689	2	chr9:35106261-35106931	637	368	Downstream
UHNhscpg0002711	0.02444	0	13	NA			NA
UHNhscpg0003020	0.02447	0.642	6	chr7:148418256-148419067	802	0	Overlapping
UHNhscpg0009776	0.0245	0	64	NA			NA
UHNhscpg0001278	0.0248	1	1	chr18:53028011-53028577	563	142141	Downstream

Sequence ID	P Values						
	Myc	Specificity Score	Total Hits	Location	Length	Gene1 Distance	Gene1 Direction
UHNhscpg0001348	0.02512	0	23	NA			NA
UHNhscpg0007695	0.02514	0.958	2	chr19:14500799-14502179	1373	0	Overlapping
UHNhscpg0000406	0.02518	1	1	chr8:71109380-71109747	364	36369	Upstream
UHNhscpg0011125	0.02525	1	1	chr7:151203104-151203495	384	1754	Upstream
UHNhscpg0000902	0.02532	1	1	chr7:45117391-45118238	844	0	Overlapping
UHNhscpg0002658	0.02533	0.704	69	chr19:21480217-21480792	480	0	Overlapping
UHNhscpg0001216	0.0254	0	42	NA			NA
UHNhscpg0000914	0.02571	0	11	NA			NA
UHNhscpg0008615	0.02583	0.03	5	chr1:40009081-40009875	795	0	Overlapping
UHNhscpg0008287	0.02595	1	1	chr9:33463227-33464212	958	0	Overlapping
UHNhscpg0005841	0.02605	1	1	chr11:33139663-33139917	255	50	Downstream
UHNhscpg0002429	0.02617	0	13	NA			NA
UHNhscpg0001688	0.02627	1	1	chr17:67859932-67860337	406	231177	Upstream
UHNhscpg0001468	0.0263	0	10	NA			NA
UHNhscpg0003078	0.02633	0.729	2	chr15:30109990-30110251	262	0	Overlapping
UHNhscpg0000097	0.02645	0.519	2	chr17:34156724-34157336	609	748	Upstream
UHNhscpg0005079	0.02653	0.86	2	chr1:152229727-152230020	257	0	Overlapping
UHNhscpg0007545	0.02667	1	1	chr11:18699693-18700388	668	3965	Upstream
UHNhscpg0000099	0.0268	0	7	chr7:74909483-74909824	338	31902	Upstream
UHNhscpg0002803	0.02687	0	13	NA			NA
UHNhscpg0002031	0.02689	1	1	chr11:60566343-60566516	174	60026	Downstream
UHNhscpg0002091	0.02724	1	1	chr11:32405917-32406464	540	7199	Upstream
UHNhscpg0002163	0.02736	0.066	8	chr17:63708274-63708410	137	47329	Downstream
UHNhscpg0001392	0.0274	1	1	chr18:32115023-32115237	190	16462	Downstream
UHNhscpg0002263	0.02765	0.296	3	chr5:168943126-168943760	615	0	Overlapping
UHNhscpg0000239	0.02789	0.219	2	chr2:128331843-128332110	260	64	Upstream
UHNhscpg0001518	0.02797	1	1	chr17:76480469-76480901	421	99388	Downstream
UHNhscpg0000537	0.02798	0	15	NA			NA
UHNhscpg0002041	0.02811	1	1	chr13:78131007-78131426	400	0	Overlapping
UHNhscpg0001049	0.02812	1	1	chr9:118445890-118446011	118	43323	Upstream
UHNhscpg0000594	0.02828	0.635	2	chr5:172588938-172589503	550	5365	Upstream
UHNhscpg0001681	0.02845	0.913	2	chr8:110725288-110725809	518	208	Downstream
UHNhscpg0000618	0.02856	1	1	chr12:30739053-30739432	380	586	Upstream
UHNhscpg0000653	0.02867	1	1	chr6:31973516-31974144	613	73	Downstream
UHNhscpg0005500	0.02886	0	13	NA			NA
UHNhscpg0007227	0.02896	1	1	chr19:50042015-50042472	458	783	Upstream
UHNhscpg0001121	0.02899	0	44	NA			NA
UHNhscpg0001828	0.02924	0.774	2	chr17:17507496-17507756	257	17755	Downstream
UHNhscpg0004425	0.02928	1	1	chr17:56828620-56828930	252	3108	Downstream
UHNhscpg0010504	0.02939	0.845	2	chr17:59577194-59577547	354	15986	Downstream
UHNhscpg0011809	0.02948	0	47	NA			NA
UHNhscpg0001051	0.02949	0	43	NA			NA
UHNhscpg0004022	0.02978	1	1	chr3:25253404-25253585	182	191172	Downstream
UHNhscpg0007487	0.02995	1	1	chr18:50049685-50050313	617	0	Overlapping
UHNhscpg0001443	0.02996	0.325	2	chr2:176842521-176842785	197	139	Upstream
UHNhscpg0007328	0.03004	0.106	3	chr1:75374734-75375251	518	8028	Upstream
UHNhscpg0000475	0.03017	1	1	chr7:99052115-99052926	621	0	Overlapping
UHNhscpg0002255	0.03042	0	13	NA			NA
UHNhscpg0009014	0.03049	0.358	6	chr11:7965150-7965517	360	0	Overlapping
UHNhscpg0001960	0.03076	1	1	chr8:101928783-101929400	602	104047	Upstream
UHNhscpg0005607	0.03079	1	1	chr18:72335447-72335597	141	195376	Downstream
UHNhscpg0007898	0.03103	1	1	chr3:23934108-23934307	153	466	Upstream
UHNhscpg0008726	0.03138	1	1	chr4:1981480-1981586	107	933	Downstream
UHNhscpg0007508	0.03139	0	12	chr16:33863882-33864008	86	5996	Upstream
UHNhscpg0000412	0.03167	0.832	2	chr19:9980882-9981159	238	988	Upstream
UHNhscpg0010798	0.03172	1	1	chr19:50703548-50703901	346	1021	Upstream
UHNhscpg0005699	0.0319	0.72	2	chr7:27226471-27226888	418	20221	Downstream

Sequence ID	P Values	Specificity Score	Total Hits	Location	Length	Gene1 Distance	Gene1 Direction
	Myc						
UHNhscpg0001274	0.03248	1	1	chr11:124241472-124242010	527	981	Upstream
UHNhscpg0003954	0.03252	0	9	NA			NA
UHNhscpg0002617	0.03266	0.494	3	chr1:46956756-46957301	542	22	Upstream
UHNhscpg0006532	0.03267	1	1	chr4:17812603-17812754	111	180129	Downstream
UHNhscpg0006831	0.03268	1	1	chr2:169743028-169743121	94	111238	Upstream
UHNhscpg0004677	0.03283	0	39	NA			NA
UHNhscpg0000722	0.03288	0	13	NA			NA
UHNhscpg0010009	0.03314	0	13	NA			NA
UHNhscpg0007124	0.0334	0.015	3	chr12:9491372-9491683	268	160081	Upstream
UHNhscpg0006319	0.03377	1	1	chr17:75392265-75393144	872	6780	Downstream
UHNhscpg0009118	0.034	1	1	chr14:36123320-36123588	262	1783	Downstream
UHNhscpg0008799	0.03402	1	1	chr2:66526261-66526432	172	10226	Upstream
UHNhscpg0007040	0.03403	0.823	2	chr20:43973192-43973549	350	644	Upstream
UHNhscpg0011745	0.03414	0	13	NA			NA
UHNhscpg0010375	0.03421	1	1	chr2:85051032-85051165	134	592	Downstream
UHNhscpg0000037	0.03437	0.04	3	chr1:144537487-144538576	1090	98405	Upstream
UHNhscpg0001509	0.03457	0.033	2	chr1:25536125-25536615	491	782	Downstream
UHNhscpg0008984	0.0346	0.024	4	chr8:57286072-57286168	85	224	Upstream
UHNhscpg0001034	0.03467	1	1	chr5:172646777-172647068	284	42044	Upstream
UHNhscpg0009228	0.03473	0.613	2	chr17:14139087-14139267	181	5963	Downstream
UHNhscpg0002720	0.03477	1	1	chr1:144106952-144107310	355	17324	Downstream
UHNhscpg0002175	0.0348	1	1	chr3:12500629-12501447	743	0	Overlapping
UHNhscpg0004193	0.035	1	1	chr11:73167868-73168609	742	7399	Downstream
UHNhscpg0006853	0.035	1	1	chr3:27731181-27731384	204	7405	Upstream
UHNhscpg0002783	0.03517	0.051	3	chr7:61416908-61416950	39	2243969	Upstream
UHNhscpg0000567	0.03527	1	1	chr6:41997067-41997790	712	125	Upstream
UHNhscpg0001759	0.03556	0.622	3	chr4:120440916-120441182	259	53	Upstream
UHNhscpg0002523	0.03578	0	13	NA			NA
UHNhscpg0003019	0.03581	1	1	chr22:30769207-30769490	276	0	Overlapping
UHNhscpg0000311	0.03604	0.952	4	chr8:104495884-104497258	1360	0	Overlapping
UHNhscpg0002250	0.03612	1	1	chr6:1342163-1342512	342	7096	Upstream
UHNhscpg0000461	0.03659	1	1	chr6:144458000-144458718	703	0	Overlapping
UHNhscpg0009957	0.03675	0	52	NA			NA
UHNhscpg0009207	0.0368	0.092	3	chr19:51611502-51611610	109	2821	Downstream
UHNhscpg0000467	0.03682	0.661	2	chr19:4059857-4060284	416	14842	Upstream
UHNhscpg0008485	0.03683	1	1	chr7:99536509-99537172	644	0	Overlapping
UHNhscpg0001559	0.03687	0.007	3	chr1:11719003-11719321	304	275	Upstream
UHNhscpg0000908	0.037	0.661	2	chr2:15649057-15649583	493	0	Overlapping
UHNhscpg0000712	0.03702	0.531	27	chr1:147490240-147490606	303	64371	Downstream
UHNhscpg0000071	0.03758	0.638	4	chr2:85517938-85518249	312	2510	Upstream
UHNhscpg0005989	0.03775	1	1	chr6:71049050-71049719	670	0	Overlapping
UHNhscpg0001170	0.03787	1	1	chr15:93894702-93894930	229	780019	Downstream
UHNhscpg0007290	0.038	0.586	2	chr5:148705064-148705229	162	0	Overlapping
UHNhscpg0000991	0.03823	0	66	NA			NA
UHNhscpg0003024	0.0387	1	1	chr19:58388353-58388523	159	0	Overlapping
UHNhscpg0000427	0.03874	1	1	chr4:128773336-128773809	462	0	Overlapping
UHNhscpg0007080	0.03902	0.8	2	chr11:44289100-44289535	420	808	Downstream
UHNhscpg0011727	0.0391	1	1	chr18:23362839-23362920	82	343560	Downstream
UHNhscpg0002688	0.03923	0.586	4	chr8:58218508-58218861	350	149551	Downstream
UHNhscpg0002619	0.03942	0	13	NA			NA
UHNhscpg0002287	0.03951	0.103	3	chr3:159306011-159306237	223	393	Upstream
UHNhscpg0002232	0.03955	1	1	chr4:41631740-41632413	630	0	Overlapping
UHNhscpg0001812	0.03963	0.728	2	chr6:107183890-107184154	265	0	Overlapping
UHNhscpg0010967	0.03976	0.657	3	chr5:132230156-132230697	522	0	Overlapping
UHNhscpg0002424	0.03977	1	1	chr4:41631740-41632413	637	0	Overlapping
UHNhscpg0002498	0.03981	1	1	chr7:29151988-29152700	689	0	Overlapping
UHNhscpg0001181	0.04024	0.908	6	chr19:49308930-49309510	573	0	Overlapping

Sequence ID	P Values	Specificity Score	Total Hits	Location	Length	Gene1 Distance	Gene1 Direction
	Myc						
UHNhscpg0007027	0.04046	1	1	chr4:133194711-133194772	62	1095147	Downstream
UHNhscpg0001446	0.0405	1	1	chr9:139550967-139551440	466	13217	Upstream
UHNhscpg0002583	0.04072	0.738	4	chr9:135272274-135273404	1098	0	Overlapping
UHNhscpg0006063	0.04078	1	1	chr5:177344333-177344745	409	11104	Upstream
UHNhscpg0009352	0.04112	0.353	2	chr6:53331731-53332080	329	9830	Downstream
UHNhscpg0010654	0.04112	1	1	chr1:113609455-113609661	207	125232	Downstream
UHNhscpg0002671	0.04115	0.701	2	chr9:35106261-35106934	662	368	Downstream
UHNhscpg0000345	0.04138	0.701	4	chr19:61710615-61711261	643	0	Overlapping
UHNhscpg0001956	0.04141	1	1	chr9:70925408-70925639	228	53269	Downstream
UHNhscpg0001599	0.04142	1	1	chr8:56919645-56920023	363	34902	Downstream
UHNhscpg0007760	0.04142	1	1	chr18:72099907-72100284	358	430689	Downstream
UHNhscpg0010529	0.04178	1	1	chr4:24625820-24625958	135	15455	Upstream
UHNhscpg0004076	0.0418	1	1	chr3:20137200-20137313	114	65374	Upstream
UHNhscpg0002171	0.04187	0.154	2	chr19:13746170-13746588	395	0	Overlapping
UHNhscpg0002431	0.04193	0.96	2	chr8:128819638-128820520	871	2141	Upstream
UHNhscpg0002925	0.04196	1	1	chr18:44615052-44615385	330	115694	Upstream
UHNhscpg0004653	0.04196	1	1	chr2:218940536-218941148	505	10714	Upstream
UHNhscpg0007377	0.0421	0	13	NA			NA
UHNhscpg0009140	0.0422	1	1	chr3:131175815-131175943	125	309	Downstream
UHNhscpg0001496	0.04229	0.588	3	chr9:83397626-83398180	514	95236	Upstream
UHNhscpg0011615	0.04245	1	1	chr2:100004565-100004600	36	83877	Upstream
UHNhscpg0011499	0.04246	0	35	NA			NA
UHNhscpg0001557	0.04292	1	1	chr2:37053-37637	573	217234	Downstream
UHNhscpg0007029	0.04293	0.648	13	chr15:76023501-76024160	644	132752	Upstream
UHNhscpg0005936	0.04297	1	1	chr21:46905649-46905872	224	25695	Upstream
UHNhscpg0001684	0.04298	1	1	chr5:126170332-126170372	41	29601	Upstream
UHNhscpg0005235	0.04298	0	35	NA			NA
UHNhscpg0008819	0.0433	0.812	2	chr21:37274394-37274976	579	9397	Upstream
UHNhscpg0006842	0.0434	0.409	2	chr2:66655720-66655966	247	139685	Upstream
UHNhscpg0000304	0.04359	0.217	30	chr6:26231712-26231909	198	202	Upstream
UHNhscpg0007532	0.04367	1	1	chr18:72099905-72100286	382	430687	Downstream
UHNhscpg0000928	0.04373	1	1	chr18:32115023-32115237	190	16462	Downstream
UHNhscpg0001767	0.04378	0	13	NA			NA
UHNhscpg0008596	0.04378	0.665	2	chr16:2954154-2954698	534	0	Overlapping
UHNhscpg0000507	0.04387	1	1	chr17:26182897-26183794	868	0	Overlapping
UHNhscpg0007226	0.04421	0	4	chr6:26829529-26829956	428	61788	Downstream
UHNhscpg0001135	0.04433	0.799	4	chr19:49220922-49221654	725	0	Overlapping
UHNhscpg0009141	0.0448	1	1	chr4:102930614-102931464	851	0	Overlapping
UHNhscpg0000090	0.04487	0.627	2	chr8:101391596-101392235	606	93	Downstream
UHNhscpg0010678	0.04496	1	1	chr5:126305083-126305146	60	88999	Upstream
UHNhscpg0000919	0.04501	1	1	chr5:140709706-140710379	650	0	Overlapping
UHNhscpg0002267	0.04507	1	1	chr7:91347862-91348424	559	0	Overlapping
UHNhscpg0004664	0.04534	0.568	2	chr1:45822221-45822486	266	0	Overlapping
UHNhscpg0008339	0.04534	1	1	chr6:43135442-43136002	553	0	Overlapping
UHNhscpg0004149	0.04545	0	34	NA			NA
UHNhscpg0001796	0.04549	0.922	2	chr3:44664988-44665757	766	0	Overlapping
UHNhscpg0002797	0.0456	0	4	chr6:26829522-26829956	431	61781	Downstream
UHNhscpg0008125	0.0456	0.741	2	chr1:51206642-51206811	170	143	Downstream
UHNhscpg0007632	0.04593	1	1	chr3:38512615-38513127	513	0	Overlapping
UHNhscpg0006103	0.04606	1	1	chr11:15628805-15629035	231	535746	Upstream
UHNhscpg0008837	0.04606	1	1	chr15:20638128-20638397	266	251	Downstream
UHNhscpg0007720	0.04609	1	1	chr1:94085305-94085670	264	31660	Upstream
UHNhscpg0004534	0.04644	0	4	chr1:148079924-148080282	347	535	Downstream
UHNhscpg0009990	0.04657	1	1	chr7:43932064-43932260	197	261	Upstream
UHNhscpg0006327	0.04683	1	1	chr2:211025765-211025896	132	23780	Upstream
UHNhscpg0007529	0.04688	0	13	NA			NA
UHNhscpg0011699	0.0469	1	1	chr15:83775792-83775966	171	50918	Upstream

Sequence ID	P Values	Specificity Score	Total Hits	Location	Length	Gene1 Distance	Gene1 Direction
	Myc						
UHNhscpg0009281	0.04694	0	4	chr1:148070730-148071003	262	0	Overlapping
UHNhscpg0011367	0.04699	1	1	chr13:67622766-67622831	58	920302	Downstream
UHNhscpg0002979	0.04706	1	1	chr17:1534808-1535275	389	0	Overlapping
UHNhscpg0003746	0.04723	1	1	chr12:55167738-55168353	616	95	Upstream
UHNhscpg0006082	0.04726	0.444	2	chr9:88752268-88752501	234	344	Downstream
UHNhscpg0006698	0.04739	0.632	4	chr8:65656638-65656998	361	1271	Upstream
UHNhscpg0001885	0.04762	1	1	chr15:99009624-99010314	651	49267	Upstream
UHNhscpg0007131	0.04773	0.746	2	chr20:43973192-43973549	287	644	Upstream
UHNhscpg0010719	0.04774	1	1	chr6:82520050-82520703	654	903	Downstream
UHNhscpg0004253	0.04794	0	35	NA			NA
UHNhscpg0010955	0.04809	0.189	41	chr1:17103856-17104232	217	16799	Downstream
UHNhscpg0002647	0.04816	0.942	2	chr4:109760868-109761763	868	0	Overlapping
UHNhscpg0007824	0.04826	0	4	chr15:41669082-41669582	501	2961	Downstream
UHNhscpg0007068	0.04836	1	1	chr1:117403485-117404038	353	433	Downstream
UHNhscpg0004050	0.04865	0	9	NA			NA
UHNhscpg0011201	0.04871	0.613	2	chr5:87992642-87993052	411	221728	Upstream
UHNhscpg0002481	0.04901	1	1	chr1:173258979-173259568	550	0	Overlapping
UHNhscpg0001494	0.04904	0.609	2	chr3:47841476-47842283	764	0	Overlapping
UHNhscpg0001520	0.04912	0.114	2	chr10:124886983-124887034	44	10593	Downstream
UHNhscpg0011013	0.04939	1	1	chr3:82083100-82083256	157	189665	Downstream
UHNhscpg0000880	0.04948	0.456	9	chr1:67668277-67668926	599	0	Overlapping
UHNhscpg0011339	0.04952	1	1	chr5:34721601-34721739	131	29327	Upstream

		Best Hit					
Gene1 Symbol	Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box
UQCRFS1	7386	1812097	Downstream	POP4	10775	0	NA
AMZ1	155185	103226	Upstream	TTYH3	80727	5	0.001
TLE1	7088	395328	Downstream	FLJ46321	389763	2	1
KIAA0090	23065	0	Overlapping	C1orf33	51154	3	0.804
TRAPPC6A	79090	259	Downstream	BLOC1S3	388552	2	0.005
METTL1	4234	329	Upstream	METTL1	4234	2	0.404
UQCRFS1	7386	1812121	Downstream	POP4	10775	0	NA
FAM90A1	55138	54582	Upstream	ZNF705A	440077	2	0
DUS3L	56931	5855	Downstream	MGC24975	163154	3	0.011
CSDA	8531	47849	Downstream	STYK1	55359	3	0.574
SMARCA2	6595	463671	Downstream	VLDLR	7436	2	0.069
ZNF225	7768	18594	Upstream	ZNF224	7767	0	NA
DAP	1611	527773	Upstream	ROPN1L	83853	1	0.006
DUS3L	56931	6454	Downstream	MGC24975	163154	2	0
ZNF557	79230	49473	Upstream	MBD3L2	125997	0	NA
SNRPB	6628	38741	Upstream	ZNF343	79175	0	NA
ZNF430	80264	61247	Downstream	ZNF714	148206	0	NA
RPL23A	6147	2003	Downstream	RAB34	83871	0	NA
KBTBD7	84078	61340	Downstream	KBTBD6	89890	0	NA
GDAP2	54834	387	Downstream	WDR3	10885	0	NA
C20orf39	79953	265312	Downstream	CST7	8530	0	NA
NA			NA	NA			NA
RLN2	6019	35390	Upstream	RLN1	6013	0	NA
LOC441426	441426	1218276	Downstream	ANKRD20A3	441425	0	NA
RPL34	6164	0	Overlapping	RPL34	6164	1	0
RPS16	6217	9456	Downstream	SUPT5H	6829	0	NA
DDX1	1653	30118	Downstream	NAG	51594	1	0
DAP	1611	527773	Upstream	ROPN1L	83853	1	0.006
SMARCA2	6595	605949	Downstream	VLDLR	7436	0	NA
SGOL1	151648	65374	Upstream	SGOL1	151648	0	NA
TMEM16E	203859	144091	Downstream	SLC17A6	57084	1	0
NA			NA	NA			NA
ZNF552	79818	29815	Downstream	ZNF587	84914	1	0.001
SERPINB9	5272	62155	Upstream	SERPINB6	5269	0	NA
HOXB2	3212	12026	Downstream	HOXB1	3211	0	NA
MANEA	79694	438545	Downstream	FUT9	10690	0	NA
LOC642265	642265	82300	Downstream	FAM74A1	401507	1	0
SIN3A	25942	85104	Downstream	MAN2C1	4123	0	NA
PTBP1	5725	22041	Upstream	PRG2	79948	0	NA
UBE2I	7329	1035	Upstream	UBE2I	7329	0	NA
BCLAF1	9774	39619	Downstream	FAM54A	113115	0	NA
TRIM26	7726	44995	Downstream	FLJ45422	441140	1	0.001
TUT1	64852	10080	Upstream	MTA2	9219	0	NA
ZNF518	9849	86987	Upstream	CCNJ	54619	0	NA
EXOD1	112479	0	Overlapping	LOC81691	81691	0	NA
INTS7	25896	0	Overlapping	DTL	51514	0	NA
UBE2D2	7322	77935	Downstream	TMEM173	340061	0	NA
GNB2L1	10399	17012	Upstream	TRIM52	84851	0	NA
ANKMY2	57037	697	Downstream	BZW2	28969	0	NA
NA			NA	NA			NA
CBX4	8535	40562	Downstream	CBX8	57332	1	1
DAP	1611	527773	Upstream	ROPN1L	83853	1	0.006
WRB	7485	31110	Downstream	HMGN1	3150	0	NA
DHX37	57647	4362	Downstream	BRI3BP	140707	1	0.062
PAK1IP1	55003	27903	Downstream	TMEM14C	51522	1	0.012

				Best Hit			
Gene1 Symbol	Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box
ZNF708	7562	37592	Downstream	ZNF493	284443	0	NA
NA			NA	NA			NA
CLASP1	23332	206290	Upstream	MKI67IP	84365	1	0.001
SERPINB9	5272	62155	Upstream	SERPINB6	5269	0	NA
NUP155	9631	0	Overlapping	NUP155	9631	0	NA
PHLPP	23239	175272	Upstream	ZCCHC2	54877	1	0
DDX11	1663	251901	Upstream	FAM60A	58516	2	0
MAP3K6	9064	8928	Upstream	FCN3	8547	0	NA
CCRN4L	25819	68863	Upstream	ELF2	1998	2	0.001
USP48	84196	385	Downstream	USP48	84196	0	NA
MDN1	23195	9913	Downstream	CASP8AP2	9994	2	0.581
RPS27L	51065	31786	Downstream	RAB8B	51762	0	NA
LOC441242	441242	102113	Downstream	VKORC1L1	154807	1	0
SGOL2	151246	16122	Downstream	KCTD18	130535	0	NA
BAI2	576	55019	Upstream	SPOCD1	90853	0	NA
RPL38	6169	9029	Downstream	TTYH2	94015	0	NA
NA			NA	NA			NA
MTX3	345778	43790	Downstream	THBS4	7060	2	0.946
ZNF222	7673	22006	Upstream	ZNF230	7773	0	NA
ELF1	1997	131747	Upstream	SLC25A15	10166	0	NA
MLL5	55904	255667	Upstream	SRPK2	6733	1	0.008
PROK2	60675	345495	Upstream	RYBP	23429	0	NA
LOC441426	441426	1218276	Downstream	ANKRD20A3	441425	0	NA
ZNF322A	79692	148399	Upstream	ABT1	29777	0	NA
TCERG1	10915	68570	Upstream	GPR151	134391	1	0.999
TEGT	7009	34650	Downstream	FMNL3	91010	0	NA
LDLR	3949	66029	Upstream	SPBC24	147841	0	NA
MDN1	23195	9909	Downstream	CASP8AP2	9994	2	0.581
CBX5	23468	413	Downstream	HNRPA1	3178	2	0.973
ERCC1	2067	1728	Downstream	ERCC1	2067	1	0.003
PTP4A2	8073	83836	Downstream	KHDRBS1	10657	0	NA
MTHFSD	64779	12212	Downstream	FOXC2	2303	3	0
SF3B5	83443	54643	Downstream	STX11	8676	1	0
ZNF35	7584	23472	Upstream	ZNF197	10168	0	NA
C11orf1	64776	7216	Downstream	ALG9	79796	2	0.001
GPATC8	23131	53781	Downstream	FZD2	2535	0	NA
CAPG	822	13381	Downstream	SH2D6	284948	0	NA
TUT1	64852	10076	Upstream	MTA2	9219	0	NA
KIAA0495	57212	4745	Downstream	CCDC27	148870	1	0
LOC441426	441426	1218276	Downstream	ANKRD20A3	441425	0	NA
PCDHGB3	56102	0	Overlapping	PCDHGB3	56102	1	0
C19orf53	28974	9834	Upstream	MGC3207	84245	1	0
PFDN5	5204	3514	Downstream	C12orf10	60314	1	0.001
LOC162993	162993	49985	Upstream	FBXL12	54850	0	NA
MID1IP1	58526	239743	Upstream	TSPAN7	7102	0	NA
NA			NA	NA			NA
PRCD	768206	19717	Downstream	CYGB	114757	0	NA
BARX1	56033	84587	Downstream	PTPDC1	138639	0	NA
POLR1B	84172	42169	Downstream	CHCHD5	84269	2	0.998
WDR45L	56270	84226	Upstream	FOXK2	3607	3	0.001
ALDH7A1	501	5438	Downstream	RNUXA	51808	1	0
UCK2	7371	59352	Downstream	TMCO1	54499	0	NA
NA			NA	NA			NA
UQCRFS1	7386	1812097	Downstream	POP4	10775	0	NA
RAX	30062	45823	Upstream	CPLX4	339302	1	0.61
COPS4	51138	24019	Downstream	DKFZp686L1814	132660	1	0
PLA2G1B	5319	14847	Upstream	SIRT4	23409	4	0.86

		Best Hit					
Gene1 Symbol	Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box
KIAA1539	80256	13107	Downstream	STOML2	30968	1	1
STAB2	55576	75173	Downstream	C12orf42	374470	0	NA
NA			NA	NA			NA
NA			NA	NA			NA
HIST2H4A	8370	0	Overlapping	H4/o	554313	0	NA
TNPO2	30000	10443	Upstream	C19orf43	79002	1	0.03
C6orf62	81688	53368	Downstream	GMNN	51053	1	0
HSBP1	3281	610006	Upstream	CDH13	1012	1	0
TBRG4	9238	45685	Downstream	RAMP3	10268	1	0
MDH1B	130752	887	Downstream	FASTKD2	22868	1	0
XRCC6	2547	550	Downstream	D15Wsu75e	27351	0	NA
NA			NA	NA			NA
FLJ40722	285966	50346	Upstream	LOC441294	441294	1	0.001
PKIB	5570	138060	Upstream	PKIB	5570	1	0
SCFD2	152579	11601	Downstream	FIP1L1	81608	2	0.976
ROBO3	64221	30941	Upstream	ROBO4	54538	1	0
ESD	2098	99301	Upstream	HTR2A	3356	1	0
SLC5A7	60482	648289	Downstream	SULT1C3	442038	0	NA
IWS1	55677	107481	Upstream	PROC	5624	0	NA
SPG20	23111	83241	Downstream	SOHLH2	54937	2	0.004
POP5	51367	0	Overlapping	POP5	51367	2	0.999
RWDD1	51389	41862	Upstream	C6orf78	221301	0	NA
KLHL18	23276	11	Downstream	KIF9	64147	1	1
TPRKB	51002	35456	Downstream	NAT8B	51471	0	NA
MEA1	4201	0	Overlapping	KLHDC3	116138	1	0.558
ZNF552	79818	29815	Downstream	ZNF587	84914	1	0.001
IARS	3376	0	Overlapping	IARS	3376	1	0
GCAT	23464	2304	Upstream	H1F0	3005	0	NA
RPS27L	51065	31786	Downstream	RAB8B	51762	0	NA
ZNF616	90317	49924	Downstream	PPP2R1A	5518	1	0
C8orf37	157657	135234	Upstream	PLEKHF2	79666	0	NA
CDC7	8317	99585	Downstream	HFM1	164045	0	NA
B3GALT2	8707	132329	Upstream	CDC73	79577	0	NA
IQGAP3	128239	19004	Downstream	APOA1BP	128240	0	NA
NA			NA	NA			NA
RPS5	6193	21454	Downstream	ZNF584	201514	0	NA
FBXO5	26271	18823	Upstream	MTRF1L	54516	2	0.022
NA			NA	NA			NA
CCDC38	120935	677	Downstream	AMDHD1	144193	1	0
ARHGAP12	94134	128113	Upstream	KIF5B	3799	0	NA
DDX1	1653	30109	Downstream	NAG	51594	1	0
DIRAS3	9077	181716	Upstream	GPR177	79971	1	0.997
IQGAP3	128239	19004	Downstream	APOA1BP	128240	0	NA
SIX3	6496	81041	Upstream	SIX2	10736	0	NA
CD3EAP	10849	1354	Downstream	PPP1R13L	10848	2	0.024
HSBP1	3281	610006	Upstream	CDH13	1012	1	0
EED	8726	56857	Downstream	C11orf73	51501	0	NA
ARL4D	379	73930	Upstream	TMEM106A	113277	0	NA
NA			NA	NA			NA
RPLP1	6176	179758	Upstream	KIF23	9493	1	0.001
MTPN	136319	228112	Downstream	UNQ1940	389558	0	NA
HD	3064	111847	Upstream	GRK4	2868	0	NA
CDC14A	8556	919	Downstream	CDC14A	8556	0	NA
ERCC1	2067	626	Downstream	ERCC1	2067	0	NA
ABCF1	23	13920	Downstream	GNL1	2794	2	0
NA			NA	NA			NA
TBC1D22A	25771	23845	Downstream	CERK	64781	1	0.004

		Best Hit					
Gene1 Symbol	**Gene1 ID**	**Gene2 Distance**	**Gene2 Direction**	**Gene2 Symbol**	**Gene2 ID**	**Total E-Boxes**	**Best E-Box**
NA			NA	NA			NA
HIST1H2BL	8340	67	Downstream	HIST1H2AI	8329	0	NA
NA			NA	NA			NA
DKFZP434O047	26083	123016	Upstream	GOSR1	9527	0	NA
OGG1	4968	183	Downstream	OGG1	4968	0	NA
IRF1	3659	53175	Upstream	IL5	3567	0	NA
GNG2	54331	128791	Downstream	C14orf166	51637	0	NA
BTBD3	22903	588787	Downstream	BTBD3	22903	0	NA
NA			NA	NA			NA
ZNF234	10780	23147	Downstream	ZNF226	7769	1	0
LOC153364	153364	270	Downstream	POLR3G	10622	4	1
NBPF1	55672	74089	Downstream	NBPF10	440673	0	NA
NA			NA	NA			NA
CHRM3	1131	704733	Downstream	FMN2	56776	0	NA
ETV1	2115	851654	Upstream	DGKB	1607	0	NA
UBE2F	140739	94093	Downstream	SCLY	51540	1	0
MKL2	57496	151537	Upstream	ERCC4	2072	0	NA
MANEA	79694	438545	Downstream	FUT9	10690	0	NA
NA			NA	NA			NA
INTS7	25896	0	Overlapping	DTL	51514	0	NA
C9orf123	90871	1041642	Upstream	JMJD2C	23081	1	0
PRDM1	639	104442	Downstream	PRDM1	639	0	NA
ZNF24	7572	32816	Upstream	ZNF396	252884	1	0.002
C14orf133	63894	23	Downstream	AHSA1	10598	0	NA
RSHL2	83861	44740	Upstream	TAGAP	117289	1	0.003
NA			NA	NA			NA
RAB37	326624	26498	Downstream	CD300E	342510	0	NA
NA			NA	NA			NA
PYDC1	260434	1208	Upstream	TRIM72	493829	1	0
NA			NA	NA			NA
SLC26A6	65010	1030	Downstream	SLC26A6	65010	1	0.001
RPS27L	51065	31786	Downstream	RAB8B	51762	0	NA
NA			NA	NA			NA
OR4D9	390199	47485	Upstream	OR4D11	219986	2	0
ESD	2098	99301	Upstream	HTR2A	3356	1	0
DHX16	8449	17813	Upstream	NRM	11270	0	NA
FLJ90650	206338	111544	Upstream	AP3S1	1176	0	NA
ZNF559	84527	18892	Downstream	ZNF699	374879	1	0
ESD	2098	99301	Upstream	HTR2A	3356	1	0
SP8	221833	130145	Upstream	ABCB5	340273	0	NA
PLCB4	5332	446539	Downstream	C20orf103	24141	0	NA
ZIC2	7546	9502	Downstream	ZIC5	85416	0	NA
DKFZp666G057	283726	65	Downstream	EFTUD1	79631	1	0
ZNF17	7565	21177	Upstream	ZNF548	147694	1	0
NA			NA	NA			NA
MGC50811	375307	13847	Downstream	SLC11A1	6556	0	NA
SMN2	6607	24377	Upstream	SERF1A	8293	0	NA
VDP	8615	50950	Downstream	G3BP2	9908	1	0.539
ATAD1	84896	45446	Downstream	PTEN	5728	0	NA
FLCN	201163	426	Upstream	FLCN	201163	0	NA
TBC1D22A	25771	23845	Downstream	CERK	64781	1	0.004
TNFAIP3	7128	239212	Upstream	PERP	64065	0	NA
GALR1	2587	102127	Downstream	MBP	4155	2	0.001
RPL21	6144	18795	Downstream	RASL11A	387496	0	NA
NA			NA	NA			NA
ZMYM1	79830	47212	Downstream	ZMYM6	9204	0	NA
POLR3A	11128	4203	Downstream	RPS24	6229	2	0.266

			Best Hit				
Gene1 Symbol	Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box
FLJ21511	80157	605777	Downstream	OCIAD2	132299	1	0
SERBP1	26135	122636	Upstream	IL12RB2	3595	0	NA
BRWD2	55717	381578	Upstream	C10orf85	404216	2	0
NEUROD2	4761	52623	Downstream	PPP1R1B	84152	1	0.001
BNC1	646	81830	Downstream	SH3GL3	6457	0	NA
SMN2	6607	24377	Upstream	SERF1A	8293	0	NA
TUT1	64852	10080	Upstream	MTA2	9219	0	NA
AP2A2	161	59315	Downstream	CHID1	66005	2	0
NA			NA	NA			NA
IQGAP3	128239	19004	Downstream	APOA1BP	128240	0	NA
TRAF7	84231	6304	Upstream	RAB26	25837	1	0.998
PGD	5226	30436	Downstream	APITD1	378708	1	0.001
COL9A1	1297	19775	Upstream	COL9A1	1297	0	NA
KBTBD6	89890	61491	Upstream	KBTBD7	84078	0	NA
KIAA0143	23167	576258	Upstream	KCNQ3	3786	1	0.006
UHMK1	127933	63544	Downstream	UAP1	6675	1	0
ATP6V1G1	9550	23138	Downstream	C9orf91	203197	1	0.179
ZNF184	7738	130338	Upstream	LOC346157	346157	0	NA
KIAA0182	23199	134821	Upstream	GINS2	51659	0	NA
ASPHD2	57168	118923	Upstream	HPS4	89781	1	0.001
CDS1	1040	84002	Downstream	NKX6-1	4825	0	NA
CLNS1A	1207	47722	Upstream	AQP11	282679	2	0
PET112L	5188	306816	Upstream	FBXW7	55294	0	NA
FXR1	8087	76325	Upstream	DNAJC19	131118	0	NA
DDX23	9416	13702	Upstream	RND1	27289	0	NA
NBPF1	55672	131670	Downstream	LOC388692	388692	0	NA
GUSBL2	375513	145821	Upstream	GUSBL2	375513	0	NA
NBPF1	55672	74089	Downstream	NBPF10	440673	0	NA
GAP43	2596	510980	Downstream	ZBTB20	26137	0	NA
RG9MTD2	93587	10627	Downstream	MTTP	4547	0	NA
SLC1A5	6510	41698	Downstream	STRN4	29888	0	NA
TUT1	64852	10080	Upstream	MTA2	9219	0	NA
CD44	960	222428	Upstream	PDHX	8050	0	NA
MPHOSPH1	9585	564951	Downstream	PANK1	53354	0	NA
SLC5A7	60482	648288	Downstream	SULT1C3	442038	0	NA
ZNF430	80264	61247	Downstream	ZNF714	148206	0	NA
MCEE	84693	52	Downstream	MPHOSPH10	10199	2	0.008
C5orf35	133383	41391	Upstream	MIER3	166968	1	0
RPL21	6144	18795	Downstream	RASL11A	387496	0	NA
MYC	4609	214454	Upstream	TMEM75	641384	1	0
ARL6IP1	23204	10734	Downstream	RPS15A	6210	0	NA
NA			NA	NA			NA
ASPHD2	57168	118923	Upstream	HPS4	89781	1	0.001
NA			NA	NA			NA
SPRYD4	283377	13831	Downstream	MIP	4284	1	0.034
FKSG30	440915	48243	Downstream	C2orf27	29798	4	0.005
LPHN2	23266	2525506	Upstream	TTLL7	79739	0	NA
C8orf37	157657	135222	Upstream	PLEKHF2	79666	0	NA
OXR1	55074	112681	Upstream	ABRA	137735	0	NA
BCAT2	587	38940	Upstream	FGF21	26291	0	NA
MYC	4609	214457	Upstream	TMEM75	641384	1	0
RPL28	6158	9071	Upstream	MDAC1	147744	0	NA
CNN1	1264	9362	Downstream	ECSIT	51295	0	NA
NGLY1	55768	7161	Downstream	OXSM	54995	1	0
LRRC33	375387	70593	Downstream	WDR53	348793	0	NA
NBPF1	55672	74089	Downstream	NBPF10	440673	0	NA
INTS7	25896	0	Overlapping	DTL	51514	0	NA

			Best Hit				
Gene1 Symbol	Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box
DDX50	79009	54306	Downstream	DDX21	9188	0	NA
LOC134145	134145	274	Downstream	CCT5	22948	0	NA
JAK2	3717	191410	Upstream	RCL1	10171	0	NA
ZC3H15	55854	103109	Downstream	ITGAV	3685	2	0.446
NA			NA	NA			NA
BCL7C	9274	21250	Downstream	CTF1	1489	1	0.266
NA			NA	NA			NA
MID1IP1	58526	239739	Upstream	TSPAN7	7102	0	NA
NPM3	10360	6818	Downstream	FGF8	2253	2	0.998
HELT	391723	128244	Downstream	SLC25A4	291	0	NA
NA			NA	NA			NA
NBPF8	641559	91643	Upstream	NBPF11	200030	1	0
ATXN2L	11273	1222	Upstream	ATXN2L	11273	0	NA
NA			NA	NA			NA
GULP1	51454	497357	Upstream	DIRC1	116093	1	0
GRAP	10750	59030	Downstream	FAM83G	644815	0	NA
RNASEN	29102	354	Downstream	C5orf22	55322	0	NA
JAG2	3714	11541	Upstream	NUDT14	256281	0	NA
ATP1A1	476	152117	Upstream	CD58	965	0	NA
C11orf1	64776	7237	Downstream	ALG9	79796	2	0.001
ZNF790	388536	11309	Downstream	ZNF345	25850	1	0.003
CXCR7	57007	246110	Upstream	CXCR7	57007	0	NA
FOXD2	2306	29172	Upstream	FOXE3	2301	0	NA
C6orf170	221322	100517	Downstream	GJA1	2697	0	NA
KIAA1008	22894	0	Overlapping	C13orf24	10464	0	NA
FLJ38723	255180	79242	Downstream	NLF2	388125	0	NA
NA			NA	NA			NA
CLDN23	137075	160249	Upstream	MFHAS1	9258	0	NA
C9orf39	54875	171875	Downstream	BNC2	54796	1	0
NA			NA	NA			NA
ZNF675	171392	71565	Upstream	ZNF681	148213	0	NA
FLJ35258	284297	31740	Upstream	NAT14	57106	0	NA
CDCA2	157313	549	Downstream	KCTD9	54793	0	NA
MTDH	92140	131707	Downstream	LAPTM4B	55353	2	0.077
CALM1	801	51034	Downstream	C14orf102	55051	0	NA
DCTD	1635	256166	Upstream	FLJ30277	152641	1	0
LOC641522	641522	65998	Upstream	LRRC37A	9884	0	NA
FIBP	9158	1840	Downstream	CCDC85B	11007	0	NA
OAS3	4940	31185	Upstream	OAS1	4938	1	0
SOX14	8403	230776	Downstream	CLDN18	51208	2	0.003
CXCR4	7852	648186	Downstream	CXCR4	7852	1	0
NA			NA	NA			NA
TRIM15	89870	8456	Upstream	TRIM15	89870	1	0
FBXO33	254170	165711	Upstream	CTAGE5	4253	1	0.728
C7orf28B	221960	176572	Upstream	LOC222967	222967	0	NA
HFE2	148738	113567	Upstream	NOTCH2NL	388677	0	NA
SORBS3	10174	9995	Upstream	SORBS3	10174	0	NA
LOC388692	388692	85603	Downstream	NBPF1	55672	0	NA
NA			NA	NA			NA
PRAMEF16	654348	38855	Upstream	PRAMEF9	343070	0	NA
DEFB131	644414	21446	Downstream	LOC650293	650293	0	NA
ZNF418	147686	12348	Upstream	ZNF256	10172	0	NA
KIAA1539	80256	13107	Downstream	STOML2	30968	1	1
NA			NA	NA			NA
ZNF786	136051	35244	Upstream	ZNF425	155054	1	0
NA			NA	NA			NA
ST8SIA3	51046	225337	Downstream	ONECUT2	9480	0	NA

			Best Hit				
Gene1 Symbol	Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box
NA			NA	NA			NA
GPSN2	9524	10598	Downstream	DNAJB1	3337	0	NA
PRDM14	63978	199618	Downstream	SLCO5A1	81796	0	NA
PRKAG2	51422	60214	Downstream	PRKAG2	51422	1	0
TBRG4	9238	45685	Downstream	RAMP3	10268	1	0
ZNF429	353088	108447	Upstream	ZNF493	284443	0	NA
NA			NA	NA			NA
NA			NA	NA			NA
OXCT2	64064	17245	Upstream	BMP8B	656	0	NA
NOL6	65083	25637	Downstream	AQP3	360	1	0.844
CSTF3	1479	50	Downstream	CSTF3	1479	0	NA
NA			NA	NA			NA
SOX9	6662	740090	Upstream	SLC39A11	201266	0	NA
NA			NA	NA			NA
CHRNA7	1139	375156	Downstream	OTUD7A	161725	1	0.041
PCGF2	7703	5191	Downstream	PSMB3	5691	0	NA
RPS27	6232	4297	Downstream	RAB13	5872	0	NA
IGSF22	283284	22767	Upstream	TMEM86A	144110	1	0.123
NSUN5B	155400	46645	Upstream	TRIM73	375593	1	1
NA			NA	NA			NA
CD5	921	70594	Upstream	CD6	923	2	0
WT1	7490	9786	Downstream	WIT1	51352	1	0
AMZ2	51321	90590	Upstream	SLC16A6	9120	0	NA
FHOD3	80206	93544	Upstream	MOCOS	55034	1	0.001
CCDC99	54908	53110	Downstream	DOCK2	1794	0	NA
POLR2D	5433	26794	Upstream	MGC4268	83607	0	NA
CHMP6	79643	142655	Downstream	BAIAP2	10458	3	0.986
NA			NA	NA			NA
C13orf7	79596	55311	Downstream	POU4F1	5457	1	1
ASTN2	23245	43418	Downstream	TRIM32	22954	0	NA
NKX2-5	1482	84784	Upstream	BNIP1	662	0	NA
FLJ20366	55638	103803	Upstream	EBAG9	9166	1	1
IPO8	10526	59283	Upstream	C1QDC1	65981	0	NA
EHMT2	10919	3604	Upstream	ZBTB12	221527	1	1
NA			NA	NA			NA
PVRL2	5819	783	Upstream	PVRL2	5819	0	NA
NA			NA	NA			NA
RAI1	10743	71777	Downstream	PEMT	10400	1	0.011
TBX2	6909	14963	Downstream	C17orf82	388407	2	0.007
ERN1	2081	15986	Downstream	ERN1	2081	2	0.999
NA			NA	NA			NA
NA			NA	NA			NA
RARB	5915	427207	Upstream	TOP2B	7155	1	0.012
POLI	11201	44529	Downstream	MBD2	8932	1	0
MTX2	10651	80969	Upstream	HOXD1	3231	0	NA
LHX8	431707	403307	Upstream	TYW3	127253	0	NA
ZNF498	221785	57732	Upstream	ZNF655	79027	0	NA
NA			NA	NA			NA
EIF3S5	8665	23515	Downstream	NALP10	338322	1	0.063
YWHAZ	7534	105345	Upstream	YWHAZ	7534	1	0.001
FLJ44881	400661	329506	Downstream	ZNF236	7776	0	NA
RPL15	6138	567	Downstream	NKIRAS1	28512	0	NA
WHSC2	7469	49798	Downstream	NAT8L	339983	0	NA
LOC649159	649159	694867	Upstream	TP53TG3	24150	0	NA
COL5A3	50509	3765	Downstream	RDH8	50700	1	0
VASP	7408	11395	Downstream	RTN2	6253	0	NA
HOXA13	3209	21800	Downstream	EVX1	2128	0	NA

Best Hit

Gene1 Symbol	Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box
ROBO3	64221	30941	Upstream	ROBO4	54538	1	0
NA			NA	NA			NA
KIAA0494	9813	50070	Downstream	ATPAF1	64756	0	NA
LCORL	254251	390981	Upstream	NCAPG	64151	0	NA
DHRS9	10170	113484	Upstream	DHRS9	10170	0	NA
NA			NA	NA			NA
NA			NA	NA			NA
NA			NA	NA			NA
KLRB1	3820	221892	Downstream	CLEC2D	29121	0	NA
CBX8	57332	25678	Upstream	CBX2	84733	1	0
NKX2-8	26257	64666	Downstream	TITF1	7080	0	NA
MEIS1	4211	951513	Downstream	ETAA1	54465	0	NA
PLTP	5360	19580	Upstream	PPGB	5476	0	NA
NA			NA	NA			NA
KCMF1	56888	64759	Upstream	TMSB10	9168	1	0.052
PDZK1	5174	110565	Downstream	CD160	11126	0	NA
TMEM50A	23585	64558	Upstream	RHD	6007	1	0.002
PLAG1	5324	700	Downstream	CHCHD7	79145	0	NA
STC2	8614	51909	Downstream	NKX2-5	1482	1	0
HS3ST3B1	9953	58212	Downstream	CDRT15	146822	0	NA
HFE2	148738	42628	Downstream	TXNIP	10628	0	NA
TSEN2	80746	72146	Downstream	MKRN2	23609	1	0.026
MRPL48	51642	18019	Downstream	RAB6A	5870	1	0
EOMES	8320	257932	Downstream	SLC4A7	9497	1	0
ZNF680	340252	2346995	Downstream	ZNF588	51427	0	NA
BYSL	705	212	Downstream	TRFP	9477	1	0.652
LOC401152	401152	21584	Upstream	FABP2	2169	0	NA
NA			NA	NA			NA
SLC5A1	6523	98729	Upstream	YWHAH	7533	0	NA
SLC25A32	81034	0	Overlapping	WDSOF1	25879	3	0.831
FOXF2	2295	84489	Upstream	FOXQ1	94234	0	NA
SF3B5	83443	54643	Downstream	STX11	8676	1	0
NA			NA	NA			NA
CCDC8	83987	55007	Upstream	FLJ10781	55228	0	NA
MAP2K2	5605	42041	Downstream	ZBTB7A	51341	1	0
AP4M1	9179	191	Upstream	MCM7	4176	1	0.987
AGTRAP	57085	44636	Upstream	C1orf187	374946	4	0.001
DDX1	1653	30152	Downstream	NAG	51594	1	0
LOC388692	388692	85603	Downstream	NBPF1	55672	0	NA
SH2D6	284948	26751	Downstream	CAPG	822	0	NA
COL9A1	1297	19775	Upstream	COL9A1	1297	0	NA
NR2F2	7026	1232752	Downstream	SPATA8	145946	0	NA
GRPEL2	134266	12572	Downstream	PCYOX1L	78991	2	0
NA			NA	NA			NA
ZNF665	79788	19516	Downstream	FLJ46385	390963	0	NA
INTU	27152	97214	Downstream	SLC25A31	83447	1	0.001
ALX4	60529	215426	Upstream	EXT2	2132	0	NA
CHST9	83539	648269	Upstream	CDH2	1000	0	NA
IMPAD1	54928	697365	Downstream	PENK	5179	0	NA
NA			NA	NA			NA
SHOX2	6474	2491	Downstream	SHOX2	6474	0	NA
TMEM33	55161	46119	Downstream	WDR21B	285429	1	0.001
RTN4IP1	84816	0	Overlapping	QRSL1	55278	0	NA
UQCRQ	27089	1780	Downstream	GDF9	2661	0	NA
TMEM33	55161	46119	Downstream	WDR21B	285429	1	0.001
CPVL	54504	47945	Downstream	CHN2	1124	0	NA
ZNF225	7768	18594	Upstream	ZNF224	7767	0	NA

		Best Hit					
Gene1 Symbol	Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box
PCDH10	57575	1095147	Downstream	PCDH10	57575	0	NA
PNPLA7	375775	14724	Downstream	MRPL41	64975	0	NA
REXO4	57109	3536	Downstream	ADAMTS13	11093	0	NA
PROP1	5626	109201	Upstream	THOC3	84321	0	NA
ELOVL5	60481	185710	Upstream	GCLC	2729	0	NA
MAGI3	260425	125517	Downstream	MAGI3	260425	0	NA
KIAA1539	80256	13107	Downstream	STOML2	30968	1	1
ZNF471	57573	30060	Downstream	ZNF667	63934	0	NA
TJP2	9414	85245	Upstream	FXN	2395	0	NA
LYN	4067	71206	Downstream	TMEM68	137695	3	0.043
FLJ44881	400661	564819	Downstream	ZNF236	7776	0	NA
LGI2	55203	34896	Downstream	DKFZp761B107	91050	0	NA
SGOL1	151648	65374	Upstream	SGOL1	151648	0	NA
C19orf53	28974	9834	Upstream	MGC3207	84245	1	0
MYC	4609	209631	Upstream	TMEM75	641384	0	NA
SMAD7	4092	295628	Upstream	KIAA0427	9811	0	NA
MGC50811	375307	13847	Downstream	SLC11A1	6556	0	NA
NA			NA	NA			NA
TRH	7200	68902	Downstream	ARVP6125	442092	0	NA
TLE1	7088	395326	Downstream	FLJ46321	389763	2	1
AFF3	3899	120869	Upstream	AFF3	3899	0	NA
NA			NA	NA			NA
ACP1	52	217258	Downstream	ACP1	52	0	NA
TBC1D2B	23102	186773	Upstream	CIB2	10518	0	NA
PRMT2	3275	56186	Downstream	S100B	6285	1	0.597
LMNB1	4001	175020	Upstream	FLJ27505	389320	0	NA
NA			NA	NA			NA
HLCS	3141	25756	Downstream	DSCR6	53820	0	NA
MEIS1	4211	821979	Downstream	ETAA1	54465	0	NA
HIST1H2BC	8347	442	Downstream	HIST1H2AC	8334	0	NA
FLJ44881	400661	564817	Downstream	ZNF236	7776	0	NA
FHOD3	80206	93544	Upstream	MOCOS	55034	1	0.001
NA			NA	NA			NA
KREMEN2	79412	4900	Downstream	PAQR4	124222	2	0
C17orf41	79915	7080	Downstream	CRLF3	51379	0	NA
ZNF322A	79692	124371	Upstream	ABT1	29777	0	NA
ZNF222	7673	22006	Upstream	ZNF230	7773	0	NA
BANK1	55024	443238	Downstream	PPP3CA	5530	0	NA
RNF19	25897	6933	Downstream	RNF19	25897	1	0.001
3-Mar	115123	164352	Upstream	LMNB1	4001	0	NA
PCDHGB1	56104	0	Overlapping	PCDHGB1	56104	3	0.135
MTERF	7978	59703	Downstream	AKAP9	10142	1	0
NASP	4678	33137	Upstream	AKR1A1	10327	0	NA
KLC4	89953	93	Upstream	KLC4	89953	0	NA
NA			NA	NA			NA
ZNF35	7584	23474	Upstream	ZNF197	10168	0	NA
ZNF322A	79692	124364	Upstream	ABT1	29777	0	NA
CDKN2C	1031	1418	Downstream	CDKN2C	1031	0	NA
ENDOGL1	9941	41822	Upstream	ACVR2B	93	0	NA
INSC	387755	538260	Upstream	INSC	387755	0	NA
NIPA1	123606	52279	Downstream	NIPA2	81614	0	NA
DNTTIP2	30836	61930	Upstream	GCLM	2730	0	NA
HIST2H3C	126961	535	Downstream	HIST2H3A	333932	0	NA
URG4	55665	311	Downstream	UBE2D4	51619	1	0.001
LANCL1	10314	103686	Downstream	CPS1	1373	0	NA
NA			NA	NA			NA
AKAP13	11214	245305	Downstream	AKAP13	11214	0	NA

		Best Hit					
Gene1 Symbol	**Gene1 ID**	**Gene2 Distance**	**Gene2 Direction**	**Gene2 Symbol**	**Gene2 ID**	**Total E-Boxes**	**Best E-Box**
HIST2H4A	8370	0	Overlapping	H4/o	554313	0	NA
PCDH9	5101	1957629	Upstream	KLHL1	57626	0	NA
PRPF8	10594	30979	Upstream	MGC14376	84981	0	NA
GLS2	27165	19110	Upstream	SPRYD4	283377	1	0.985
GAS1	2619	200877	Downstream	FLJ45537	401535	0	NA
BHLHB5	27319	216904	Upstream	CYP7B1	9420	1	1
ASB7	140460	49267	Upstream	ASB7	140460	0	NA
PLTP	5360	19580	Upstream	PPGB	5476	0	NA
FAM46A	55603	493464	Upstream	IBTK	25998	0	NA
NA			NA	NA			NA
CROCC	9696	33830	Upstream	LOC440570	440570	0	NA
RPL34	6164	0	Overlapping	RPL34	6164	1	0
CKMT1B	1159	4699	Downstream	HISPPD2A	9677	0	NA
TTF2	8458	57529	Upstream	IGSF2	9398	0	NA
NA			NA	NA			NA
MEF2C	4208	392221	Downstream	TMEM161B	153396	1	0.615
MRPS14	63931	23465	Upstream	CACYBP	27101	1	0
DHX30	22907	21822	Upstream	DHX30	22907	2	0
HMX2	3167	16825	Downstream	BUB3	9184	0	NA
GBE1	2632	2361349	Downstream	ROBO1	6091	0	NA
SERBP1	26135	122643	Upstream	IL12RB2	3595	0	NA
RAI14	26064	153286	Downstream	FLJ25439	153657	0	NA

Best Location	Location	Length	Gene1 Distance	Gene1 Direction	Gene1 Symbol
NA	chr7:62451813-62452158	136	1208761	Upstream	ZNF680
chr7:2742139-2742145	chr7:2740920-2741293	355	55232	Upstream	AMZ1
chr9:83397844-83397850	chr9:81523458-81523717	212	146761	Upstream	TLE4
chr1:19450900-19450906	NA	NA	NA	NA	NA
chr19:50373795-50373801	NA	NA	NA	NA	NA
chr12:56452164-56452170	chrX:25123136-25123270	76	179361	Downstream	ARX
NA	chr7:62451813-62452158	136	1208761	Upstream	ZNF680
chr12:8271408-8271414	chr4:3990600-3991350	604	252507	Upstream	ADRA2C
chr19:5741041-5741047	NA	NA	NA	NA	NA
chr12:10766310-10766316	NA	NA	NA	NA	NA
chr9:2147925-2147931	NA	NA	NA	NA	NA
NA	chr19:49290248-49290336	53	0	Overlapping	ZNF224
chr5:11022370-11022376	NA	NA	NA	NA	NA
chr19:5741843-5741849	NA	NA	NA	NA	NA
NA	chr19:7050625-7050788	156	30155	Upstream	ZNF557
NA	NA	NA	NA	NA	NA
NA	chr19:21116591-21116939	201	0	Overlapping	ZNF431
NA	chr17:56501999-56502038	36	330000	Downstream	TBX2
NA	chr13:40604532-40605075	282	0	Overlapping	KBTBD6
NA	chr1:118273592-118274037	446	0	Overlapping	GDAP2
NA	chr20:24612659-24612734	76	214825	Upstream	C20orf39
NA	chr1:77209567-77209688	82	103794	Upstream	ST6GALNAC5
NA	chr9:5328975-5329649	404	224	Upstream	RLN1
NA	chr9:42009143-42009515	345	64112	Downstream	MGC21881
chr4:109760420-109760426	chr6:43403741-43403841	50	19233	Downstream	CRIP3
NA	chr16:2322611-2322698	52	8050	Upstream	ABCA3
chr2:15649792-15649798	NA	NA	NA	NA	NA
chr5:11022370-11022376	NA	NA	NA	NA	NA
NA	chr9:2004805-2004876	68	465	Downstream	SMARCA2
NA	NA	NA	NA	NA	NA
chr11:22172072-22172078	NA	NA	NA	NA	NA
NA	chr5:134286935-134287591	462	18227	Upstream	PCBD2
chr19:63023020-63023026	chr19:63119403-63119709	196	81	Upstream	ZNF417
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr6:96132093-96132185	85	0	Overlapping	MANEA
chr9:39807753-39807759	chr9:39277958-39278790	691	0	Overlapping	CNTNAP3
NA	NA	NA	NA	NA	NA
NA	chr19:749887-749946	60	1477	Upstream	PTBP1
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr6:30290119-30290125	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr16:20725792-20726021	222	459	Upstream	LOC81691
NA	NA	NA	NA	NA	NA
NA	chr14:18862857-18862929	45	108076	Upstream	LOC440157
NA	chr5:108147697-108147802	51	36276	Upstream	FER
NA	chr7:16651637-16651766	126	157	Upstream	ANKMY2
NA	chr1:77209567-77209688	82	103794	Upstream	ST6GALNAC5
chr17:75426285-75426291	chr17:75425774-75425980	207	1828	Upstream	CBX4
chr5:11022370-11022376	NA	NA	NA	NA	NA
NA	chr4:113834303-113834378	40	56243	Upstream	LARP7
chr12:124039584-124039590	NA	NA	NA	NA	NA
chr6:10803190-10803196	chr6:10803065-10803122	58	51	Downstream	PAK1IP1

Best Location	Location	Length	Gene1 Distance	Gene1 Direction	Gene1 Symbol
NA	chr19:21303463-21304001	302	51	Upstream	ZNF708
NA	chr1:556638-557718	980	54179	Upstream	OR4F3
chr2:122005139-122005145	chr2:122003874-122004120	240	119402	Upstream	CLASP1
NA	NA	NA	NA	NA	NA
NA	chr5:37407266-37407305	40	312	Downstream	NUP155
chr18:58532681-58532687	NA	NA	NA	NA	NA
chr12:31117703-31117709	chr12:9491697-9492229	465	159535	Upstream	KLRB1
NA	chr1:27565075-27565220	130	704	Upstream	MAP3K6
chr4:140156057-140156063	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr6:90586202-90586208	NA	NA	NA	NA	NA
NA	chr15:61237072-61237132	61	278	Downstream	RPS27L
chr7:64872879-64872885	chr7:64178589-64179449	746	89740	Downstream	H-plk
NA	NA	NA	NA	NA	NA
NA	chr1:31998486-31998769	284	3466	Upstream	BAI2
NA	NA	NA	NA	NA	NA
NA	chr5:134289001-134289707	526	20293	Upstream	PCBD2
chr5:79322218-79322224	NA	NA	NA	NA	NA
NA	chr19:49222671-49222844	146	1316	Upstream	ZNF222
NA	chr9:99040526-99040788	165	0	Overlapping	KIAA1529
chr7:104441533-104441539	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr9:42009143-42009506	340	64112	Downstream	MGC21881
NA	chr6:58256529-58256963	387	138692	Upstream	GUSBL2
chr5:145806777-145806783	NA	NA	NA	NA	NA
NA	chr12:48421533-48421869	337	10642	Downstream	TEGT
NA	NA	NA	NA	NA	NA
chr6:90586202-90586208	NA	NA	NA	NA	NA
chr12:52960312-52960318	NA	NA	NA	NA	NA
chr19:50620785-50620791	chr19:50619274-50619314	41	257	Downstream	ERCC1
NA	chr1:32168543-32168581	35	7994	Upstream	PTP4A2
chr16:85146022-85146028	chr16:85146282-85146478	197	0	Overlapping	MTHFSD
chr6:144458684-144458690	NA	NA	NA	NA	NA
NA	chr5:42504311-42504406	60	44529	Upstream	GHR
chr11:111254805-111254811	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr1:3654470-3654476	chr1:3653774-3653883	110	724	Downstream	KIAA0495
NA	chr9:42009143-42009515	345	64112	Downstream	MGC21881
chr5:140729991-140729997	NA	NA	NA	NA	NA
chr19:13746443-13746449	chr19:13746645-13746990	321	389	Upstream	C19orf53
chr12:51975314-51975320	chr1:220881673-220881762	70	51795	Downstream	TAF1A
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr5:134289000-134289705	513	20292	Upstream	PCBD2
NA	chr9:105007050-105007120	59	209567	Upstream	CYLC2
NA	chr9:95748104-95748175	51	9248	Upstream	BARX1
chr2:113016666-113016672	chr2:113015806-113015884	79	199	Downstream	POLR1B
chr17:78154963-78154969	chr7:113020280-113020364	41	325936	Upstream	PPP1R3A
chr5:125958757-125958763	chr5:76621765-76621958	158	79304	Upstream	PDE8B
NA	NA	NA	NA	NA	NA
NA	chr1:558006-558507	435	53390	Upstream	OR4F3
NA	chr7:62451813-62452008	66	1208911	Upstream	ZNF680
chr18:55090066-55090072	chr19:3722535-3722674	60	545	Upstream	RAXL1
chr4:84174913-84174919	chrX:41891087-41891182	52	223875	Downstream	CASK
chr12:119240006-119240012	NA	NA	NA	NA	NA

Best Location	Location	Length	Gene1 Distance	Gene1 Direction	Gene1 Symbol
chr9:35106373-35106379	chr9:35105993-35106202	198	100	Downstream	KIAA1539
NA	NA	NA	NA	NA	NA
NA	chr11:10486010-10486180	119	33170	Upstream	RNF141
NA	chr1:558000-558507	476	53390	Upstream	OR4F3
NA	chr1:148099191-148099464	266	0	Overlapping	HIST2H4A
chr19:12695765-12695771	NA	NA	NA	NA	NA
chr6:24829836-24829842	NA	NA	NA	NA	NA
chr16:81828081-81828087	NA	NA	NA	NA	NA
chr7:45117713-45117719	NA	NA	NA	NA	NA
chr2:207338840-207338846	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr5:79982889-79983133	217	2916	Downstream	MSH3
chr7:143029527-143029533	chr7:143163227-143163544	94	66561	Upstream	KIAA0738
chr6:122973093-122973099	NA	NA	NA	NA	NA
chr4:53926970-53926976	NA	NA	NA	NA	NA
chr11:124241626-124241632	NA	NA	NA	NA	NA
chr13:46268725-46268731	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr2:128000303-128000376	74	138	Upstream	IWS1
chr13:35769606-35769612	NA	NA	NA	NA	NA
chr12:119503341-119503347	NA	NA	NA	NA	NA
NA	chr12:109028554-109028643	62	17921	Downstream	IFT81
chr3:47299496-47299502	chr3:47298259-47298536	278	506	Upstream	KIF9
NA	chr2:73817097-73817166	70	855	Upstream	TPRKB
chr6:43089920-43089926	NA	NA	NA	NA	NA
chr19:63023020-63023026	chr19:63119403-63119709	196	81	Upstream	ZNF417
chr9:94095981-94095987	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr15:61237072-61237132	61	278	Downstream	RPS27L
chr19:57334896-57334902	chr19:57335199-57335317	119	196	Downstream	ZNF616
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr1:77209578-77209688	79	103805	Upstream	ST6GALNAC5
NA	NA	NA	NA	NA	NA
chr6:153346136-153346142	NA	NA	NA	NA	NA
NA	chr17:21944883-21945047	97	195387	Upstream	FAM27L
chr12:94860961-94860967	chr12:94860887-94860997	107	204	Downstream	AMDHD1
NA	chr10:32257290-32257349	60	427	Upstream	ARHGAP12
chr2:15649792-15649798	NA	NA	NA	NA	NA
chr1:68288176-68288182	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr2:45009107-45009145	39	13395	Downstream	SIX3
chr19:50601730-50601736	NA	NA	NA	NA	NA
chr16:81828081-81828087	NA	NA	NA	NA	NA
NA	chr11:85633611-85633697	87	149	Upstream	EED
NA	chr17:38795701-38795885	149	35993	Downstream	ARL4D
NA	chr5:134286935-134287591	462	18227	Upstream	PCBD2
chr15:67673631-67673637	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr4:3047188-3047231	44	983	Upstream	HD
NA	NA	NA	NA	NA	NA
NA	chr19:50620370-50620427	42	1353	Downstream	ERCC1
chr6:30647119-30647125	chr6:30646685-30646831	123	317	Downstream	ABCF1
NA	chr11:10486010-10486178	117	33172	Upstream	RNF141
chr22:45537300-45537306	NA	NA	NA	NA	NA

Best Location	Location	Length	Gene1 Distance	Gene1 Direction	Gene1 Symbol
NA	chr17:21944883-21945047	97	195387	Upstream	FAM27L
NA	chr6:27914419-27914708	222	1	Upstream	HIST1H2BN
NA	chr5:134290508-134290913	342	21800	Upstream	PCBD2
NA	chr17:25952007-25952238	220	41298	Upstream	DKFZP434O047
NA	chr3:9766593-9766825	233	0	Overlapping	OGG1
NA	chr5:131854039-131854258	220	68	Upstream	IRF1
NA	chr14:51397183-51397290	108	16769	Downstream	GNG2
NA	NA	NA	NA	NA	NA
NA	chr1:558000-558507	468	53390	Upstream	OR4F3
chr19:49338358-49338364	chr19:49290600-49290694	64	264	Upstream	ZNF224
chr5:89805563-89805569	chr18:44258665-44258883	127	60541	Downstream	KIAA0427
NA	chr1:16939839-16940263	377	127270	Downstream	NBPF1
NA	NA	NA	NA	NA	NA
NA	chr1:237616196-237616311	116	242684	Downstream	CHRM3
NA	NA	NA	NA	NA	NA
chr2:238540562-238540568	NA	NA	NA	NA	NA
NA	chr16:14073702-14073809	96	1006	Upstream	MKL2
NA	chr6:96132093-96132185	85	0	Overlapping	MANEA
NA	chr1:556651-557716	953	54181	Upstream	OR4F3
NA	NA	NA	NA	NA	NA
chr9:7789387-7789393	chr15:22965642-22965934	167	214415	Upstream	SNRPN
NA	NA	NA	NA	NA	NA
chr18:31178648-31178654	chr18:31177573-31177812	240	591	Upstream	ZNF24
NA	NA	NA	NA	NA	NA
chr6:159341088-159341094	chr6:159343965-159344039	39	2779	Downstream	RSHL2
NA	chr5:134286943-134287591	454	18235	Upstream	PCBD2
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr16:31134763-31134769	NA	NA	NA	NA	NA
NA	chr11:10486010-10486180	119	33170	Upstream	RNF141
chr3:48647002-48647008	NA	NA	NA	NA	NA
NA	chr15:61237072-61237132	61	278	Downstream	RPS27L
NA	chr1:558006-558507	462	53390	Upstream	OR4F3
chr11:59074846-59074852	NA	NA	NA	NA	NA
chr13:46268725-46268731	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr19:9296007-9296013	NA	NA	NA	NA	NA
chr13:46268725-46268731	NA	NA	NA	NA	NA
NA	chr7:20784444-20784799	344	8231	Upstream	SP8
NA	NA	NA	NA	NA	NA
NA	chr13:99431429-99431568	140	751	Downstream	ZIC2
chr15:80342386-80342392	chr15:82539418-82539905	280	217498	Upstream	KIAA1920
chr19:62614755-62614761	chr19:62638587-62638892	126	7697	Downstream	ZNF749
NA	chr17:21948201-21948643	143	198705	Upstream	FAM27L
NA	NA	NA	NA	NA	NA
NA	chr5:70256646-70257025	376	123	Upstream	SMN1
chr4:76868646-76868652	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr22:45537300-45537306	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr18:73075966-73075972	NA	NA	NA	NA	NA
NA	chr13:26723690-26723856	163	0	Overlapping	RPL21
NA	NA	NA	NA	NA	NA
NA	chr1:35317056-35317207	116	351	Downstream	ZMYM1
chr10:79459309-79459315	chr10:79459401-79459459	59	136	Downstream	POLR3A

Best Location	Location	Length	Gene1 Distance	Gene1 Direction	Gene1 Symbol
chr4:49209190-49209196	chr1:142124430-142124689	208	950914	Upstream	PPIAL4
NA	chr4:79321923-79322319	357	123804	Upstream	FRAS1
chr10:122728535-122728541	NA	NA	NA	NA	NA
chr17:34983952-34983958	chr17:34983307-34983397	91	34304	Upstream	NEUROD2
NA	NA	NA	NA	NA	NA
NA	chr5:70256646-70257025	372	123	Upstream	SMN1
NA	NA	NA	NA	NA	NA
chr11:960424-960430	chr11:959953-960074	102	44113	Upstream	AP2A2
NA	chr1:554327-554724	386	57173	Upstream	OR4F3
NA	NA	NA	NA	NA	NA
chr16:2145055-2145061	chr10:128458275-128458329	35	125683	Downstream	DOCK1
chr1:10381309-10381315	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr13:40666278-40666822	290	0	Overlapping	KBTBD7
chr8:132985062-132985068	chr8:132985248-132985467	216	35862	Downstream	KIAA0143
chr1:160734241-160734247	NA	NA	NA	NA	NA
chr9:116389812-116389818	chr14:105323146-105323287	50	259149	Upstream	TMEM121
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr22:25086936-25086942	chr22:25086887-25087002	116	68299	Downstream	ASPHD2
NA	NA	NA	NA	NA	NA
chr11:77025630-77025636	chr4:170990265-170990503	97	74642	Downstream	C4orf27
NA	chr4:153186090-153186280	170	284494	Downstream	PET112L
NA	NA	NA	NA	NA	NA
NA	chr12:47532233-47532522	290	9	Downstream	DDX23
NA	chr1:144656895-144657026	68	111953	Upstream	NBPF8
NA	chr6:58249909-58250236	324	145419	Upstream	GUSBL2
NA	chr1:16939839-16940263	377	127270	Downstream	NBPF1
NA	chr3:116859547-116859779	217	34406	Upstream	GAP43
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr11:35117719-35118002	273	727	Upstream	CD44
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr19:21116591-21116976	228	0	Overlapping	ZNF431
chr2:71210309-71210315	chr2:71210964-71211552	565	13	Upstream	MPHOSPH10
chr5:56241865-56241871	NA	NA	NA	NA	NA
NA	chr13:26723690-26723856	163	0	Overlapping	RPL21
chr8:128816091-128816097	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr17:21944883-21945032	90	195387	Upstream	FAM27L
chr22:25086936-25086942	chr22:25086887-25087002	116	68299	Downstream	ASPHD2
NA	chr1:556116-557716	1484	54181	Upstream	OR4F3
chr12:55148567-55148573	NA	NA	NA	NA	NA
chr2:132146992-132146998	chr14:18711904-18712927	808	41854	Downstream	LOC440157
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr8:128816091-128816097	NA	NA	NA	NA	NA
NA	chr3:171854552-171854661	70	68000	Downstream	SLC7A14
NA	NA	NA	NA	NA	NA
chr3:25798739-25798745	NA	NA	NA	NA	NA
NA	chr3:197850925-197851375	443	0	Overlapping	LRRC33
NA	chr1:16939839-16940263	377	127270	Downstream	NBPF1
NA	NA	NA	NA	NA	NA

Best Location	Location	Length	Gene1 Distance	Gene1 Direction	Gene1 Symbol
NA	chr2:32280092-32280255	116	35656	Upstream	SLC30A6
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr2:187059526-187059532	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr16:30794542-30794548	chr16:30794200-30794254	55	18646	Upstream	BCL7C
NA	chr1:558000-558507	456	53390	Upstream	OR4F3
NA	NA	NA	NA	NA	NA
chr10:103532601-103532607	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr1:145999056-145999062	chr1:144702006-144702550	537	66429	Upstream	NBPF8
NA	chr12:52432046-52432168	111	24560	Downstream	CALCOCO1
NA	chr1:558000-558500	465	53397	Upstream	OR4F3
chr2:188864671-188864677	chr2:188864724-188864924	201	710	Downstream	GULP1
NA	chr17:18956315-18956457	131	65254	Downstream	GRAP
NA	NA	NA	NA	NA	NA
NA	chr14:104706859-104706930	40	653	Downstream	JAG2
NA	chr1:116762410-116762575	166	45052	Upstream	ATP1A1
chr11:111254805-111254811	NA	NA	NA	NA	NA
chr19:42021172-42021178	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr6:121697100-121697222	123	121	Upstream	C6orf170
NA	NA	NA	NA	NA	NA
NA	chr10:30695507-30695853	163	17253	Downstream	PAPD1
NA	chr1:558000-558507	468	53390	Upstream	OR4F3
NA	NA	NA	NA	NA	NA
chr9:17032825-17032831	chr9:17032364-17032500	133	92537	Downstream	C9orf39
NA	chr5:134286935-134287591	465	18227	Upstream	PCBD2
NA	chr19:21116588-21117166	272	0	Overlapping	ZNF431
NA	chr19:60719936-60720068	133	1776	Downstream	FLJ35258
NA	NA	NA	NA	NA	NA
chr8:98725543-98725549	chr8:98725301-98725388	88	194	Downstream	MTDH
NA	NA	NA	NA	NA	NA
chr4:184001108-184001114	NA	NA	NA	NA	NA
NA	chr17:42011770-42012558	769	0	Overlapping	ARL17P1
NA	chr11:65412670-65413016	264	84	Downstream	FIBP
chr12:111860995-111861001	chr22:16640527-16640826	288	3269	Downstream	BID
chr3:138968943-138968949	NA	NA	NA	NA	NA
chr2:137238564-137238570	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr6:30247515-30247521	NA	NA	NA	NA	NA
chr14:38971494-38971500	NA	NA	NA	NA	NA
NA	chr11:67316659-67317125	403	111398	Downstream	ALDH3B2
NA	chr1:144029319-144030296	951	94338	Downstream	HFE2
NA	chr8:22475009-22475058	46	4101	Downstream	SORBS3
NA	chr1:142464219-142464519	142	611084	Upstream	PPIAL4
NA	chr1:558000-558507	468	53390	Upstream	OR4F3
NA	chr1:13553453-13553544	92	38894	Upstream	PRAMEF9
NA	chr3:75728395-75728440	42	1075362	Downstream	CNTN3
NA	chr19:63017577-63017923	142	170	Upstream	ZNF552
chr9:35106373-35106379	chr9:35105993-35106202	198	100	Downstream	KIAA1539
NA	chr1:558000-558507	476	53390	Upstream	OR4F3
chr7:148418543-148418549	chr7:148419322-148419608	287	602	Downstream	ZNF786
NA	chr1:556656-557716	900	54181	Upstream	OR4F3
NA	NA	NA	NA	NA	NA

Best Location	Location	Length	Gene1 Distance	Gene1 Direction	Gene1 Symbol
NA	NA	NA	NA	NA	NA
NA	chr4:125604379-125604459	57	246923	Upstream	ANKRD50
NA	NA	NA	NA	NA	NA
chr7:151203334-151203340	NA	NA	NA	NA	NA
chr7:45117713-45117719	NA	NA	NA	NA	NA
NA	chr19:21116627-21117208	142	0	Overlapping	ZNF431
NA	chr5:134286946-134287591	384	18238	Upstream	PCBD2
NA	chr1:558000-558507	456	53390	Upstream	OR4F3
NA	chr1:39752855-39753648	771	22951	Upstream	BMP8A
chr9:33463904-33463910	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr1:558006-558507	470	53390	Upstream	OR4F3
NA	NA	NA	NA	NA	NA
NA	chr1:558130-558507	338	53390	Upstream	OR4F3
chr15:30109929-30109935	chr15:30110318-30110388	71	301	Upstream	CHRNA7
NA	chr17:34157430-34157722	293	362	Upstream	PCGF2
NA	chr12:3191339-3191378	36	134522	Upstream	TSPAN9
chr11:18699664-18699670	NA	NA	NA	NA	NA
chr7:74909969-74909975	chr7:76094085-76094426	338	130	Upstream	POMZP3
NA	chr1:558006-558507	470	53390	Upstream	OR4F3
chr11:60565869-60565875	NA	NA	NA	NA	NA
chr11:32406378-32406384	NA	NA	NA	NA	NA
NA	chr17:60203895-60204054	128	115047	Downstream	SMURF2
chr18:32115619-32115625	NA	NA	NA	NA	NA
NA	chr1:159390194-159390634	433	22	Upstream	UFC1
NA	chr2:128332289-128332495	203	115	Downstream	POLR2D
chr17:76481158-76481164	NA	NA	NA	NA	NA
NA	chr1:558347-558507	153	53390	Upstream	OR4F3
chr13:78131204-78131210	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr5:172589588-172589792	201	5076	Upstream	NKX2-5
chr8:110726265-110726271	chr8:110725074-110725118	45	0	Overlapping	FLJ20366
NA	NA	NA	NA	NA	NA
chr6:31973479-31973485	NA	NA	NA	NA	NA
NA	chr1:558000-558507	476	53390	Upstream	OR4F3
NA	NA	NA	NA	NA	NA
NA	chr5:134286935-134287597	463	18227	Upstream	PCBD2
chr17:17508017-17508023	chr17:17507216-17507273	58	18238	Downstream	RAI1
chr17:56828136-56828142	NA	NA	NA	NA	NA
chr17:59577199-59577205	chr4:185773446-185773520	55	34103	Upstream	CASP3
NA	chr1:554327-554724	386	57173	Upstream	OR4F3
NA	chr5:134286935-134287591	469	18227	Upstream	PCBD2
chr3:25253515-25253521	NA	NA	NA	NA	NA
chr18:50049746-50049752	NA	NA	NA	NA	NA
NA	chr2:176842851-176843034	133	469	Upstream	MTX2
NA	chr1:75375433-75375899	463	8727	Upstream	LHX8
NA	NA	NA	NA	NA	NA
NA	chr1:558000-558507	468	53390	Upstream	OR4F3
chr11:7965392-7965398	chr11:7965622-7965856	231	180	Upstream	EIF3S5
chr8:101928959-101928965	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr2:132742653-132742783	86	148302	Downstream	GPR39
chr19:9980610-9980616	chr19:9980788-9980827	40	1320	Upstream	COL5A3
NA	NA	NA	NA	NA	NA
NA	chr7:27226912-27227028	117	20662	Downstream	HOXA13

Best Location	Location	Length	Gene1 Distance	Gene1 Direction	Gene1 Symbol
chr11:124241626-124241632	NA	NA	NA	NA	NA
NA	chr1:77209567-77209688	82	103794	Upstream	ST6GALNAC5
NA	chr1:35358967-35359457	274	41409	Upstream	ZMYM1
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr5:79982889-79983133	217	2916	Downstream	MSH3
NA	chr1:558000-558507	468	53390	Upstream	OR4F3
NA	chr1:558000-558507	476	53390	Upstream	OR4F3
NA	chr12:9327921-9328232	264	75688	Downstream	PZP
chr17:75392578-75392584	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr20:43973074-43973139	62	1054	Upstream	PLTP
NA	chr1:558000-558505	474	53392	Upstream	OR4F3
chr2:85050802-85050808	NA	NA	NA	NA	NA
NA	chr1:145866984-145868073	1046	0	Overlapping	GPR89A
chr1:25536197-25536203	chr1:25465796-25466286	475	5281	Downstream	RHD
NA	chr8:57285797-57285879	83	513	Upstream	PLAG1
chr5:172647030-172647036	NA	NA	NA	NA	NA
NA	chr17:14141514-14141603	70	3627	Downstream	HS3ST3B1
NA	NA	NA	NA	NA	NA
chr3:12500700-12500706	NA	NA	NA	NA	NA
chr11:73168008-73168014	NA	NA	NA	NA	NA
chr3:27731747-27731753	NA	NA	NA	NA	NA
NA	chr10:41981232-41981272	37	472726	Upstream	ZNF33B
chr6:41997243-41997249	NA	NA	NA	NA	NA
NA	chr4:120440701-120440802	98	433	Upstream	LOC401152
NA	chr1:558000-558507	476	53390	Upstream	OR4F3
NA	NA	NA	NA	NA	NA
chr8:104496740-104496746	chr15:49024119-49024219	65	35882	Upstream	AP4E1
NA	NA	NA	NA	NA	NA
chr6:144458684-144458690	NA	NA	NA	NA	NA
NA	chr1:556840-557719	785	54178	Upstream	OR4F3
NA	chr19:51611826-51611949	99	3145	Downstream	CCDC8
chr19:4059570-4059576	chr19:4059678-4059818	141	15308	Upstream	MAP2K2
chr7:99536976-99536982	NA	NA	NA	NA	NA
chr1:11718715-11718721	chr1:11718450-11718755	302	0	Overlapping	AGTRAP
chr2:15649792-15649798	chr16:28743136-28743415	167	1222	Upstream	ATXN2L
NA	chr1:142464219-142464519	142	611084	Upstream	PPIAL4
NA	chr2:85518339-85518451	113	2911	Upstream	SH2D6
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr5:148704707-148704713	chr5:148704700-148704766	67	403	Downstream	GRPEL2
NA	chr5:134289000-134289705	513	20292	Upstream	PCBD2
NA	NA	NA	NA	NA	NA
chr4:128773840-128773846	NA	NA	NA	NA	NA
NA	chr11:44289643-44289734	84	1351	Downstream	ALX4
NA	NA	NA	NA	NA	NA
NA	chr8:58218334-58218478	145	149377	Downstream	IMPAD1
NA	chr1:558000-558507	476	53390	Upstream	OR4F3
NA	chr3:159305720-159305923	200	707	Upstream	SHOX2
chr4:41632661-41632667	NA	NA	NA	NA	NA
NA	chr12:75055075-75055174	72	211179	Upstream	BBS10
NA	chr5:132229644-132229822	179	395	Downstream	UQCRQ
chr4:41632661-41632667	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr19:49290248-49290336	53	0	Overlapping	ZNF224

Best Location	Location	Length	Gene1 Distance	Gene1 Direction	Gene1 Symbol
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr15:29533409-29533753	288	127035	Upstream	KLF13
NA	NA	NA	NA	NA	NA
NA	chr6:53331361-53331573	213	9460	Downstream	ELOVL5
NA	NA	NA	NA	NA	NA
chr9:35106373-35106379	chr9:35105993-35106202	198	100	Downstream	KIAA1539
NA	chr19:61711435-61711643	192	412	Upstream	ZNF471
NA	NA	NA	NA	NA	NA
chr8:56919663-56919669	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr19:13746443-13746449	chr19:13746645-13746982	334	389	Upstream	C19orf53
NA	chr1:40139584-40139626	35	648	Upstream	MYCL1
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr1:558007-558507	469	53390	Upstream	OR4F3
NA	NA	NA	NA	NA	NA
chr9:83397844-83397850	chr9:81523458-81523717	212	146761	Upstream	TLE4
NA	NA	NA	NA	NA	NA
NA	chr11:10486010-10486181	120	33169	Upstream	RNF141
NA	NA	NA	NA	NA	NA
NA	chr15:72179475-72179803	227	17531	Downstream	GOLGA6
chr21:46905251-46905257	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr11:10486010-10486180	119	33170	Upstream	RNF141
NA	chr21:37275040-37275148	109	9225	Upstream	HLCS
NA	chr2:66656086-66656231	146	140051	Upstream	MEIS1
NA	chr6:26266579-26266757	155	252	Upstream	HIST1H2BD
NA	NA	NA	NA	NA	NA
chr18:32115619-32115625	NA	NA	NA	NA	NA
NA	chr1:558000-558507	476	53390	Upstream	OR4F3
chr16:2953823-2953829	chr16:2953917-2954095	179	122	Downstream	KREMEN2
NA	NA	NA	NA	NA	NA
NA	chr6:26888128-26888555	428	120387	Downstream	ZNF322A
NA	chr19:49222671-49222844	146	1316	Upstream	ZNF222
NA	NA	NA	NA	NA	NA
chr8:101392063-101392069	chr8:101391276-101391501	226	2	Upstream	RNF19
NA	NA	NA	NA	NA	NA
chr5:140710307-140710313	NA	NA	NA	NA	NA
chr7:91348260-91348266	NA	NA	NA	NA	NA
NA	chr8:62014926-62015084	115	261034	Upstream	CHD7
NA	NA	NA	NA	NA	NA
NA	chr11:10486010-10486180	119	33170	Upstream	RNF141
NA	chr5:42504311-42504406	60	44529	Upstream	GHR
NA	chr6:26888128-26888562	431	120387	Downstream	ZNF322A
NA	chr1:51206925-51206968	44	0	Overlapping	CDKN2C
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr1:148089912-148090270	347	534	Downstream	HIST2H3C
chr7:43932122-43932128	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
NA	chr1:558000-558507	476	53390	Upstream	OR4F3
NA	NA	NA	NA	NA	NA

Best Location	Location	Length	Gene1 Distance	Gene1 Direction	Gene1 Symbol
NA	chr1:148099191-148099464	262	0	Overlapping	HIST2H4A
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
chr12:55167696-55167702	NA	NA	NA	NA	NA
NA	chr9:88752519-88752648	130	595	Downstream	GAS1
chr8:65657029-65657035	chr20:61108007-61108187	133	600	Upstream	BHLHB4
NA	NA	NA	NA	NA	NA
NA	chr20:43973066-43973138	73	1055	Upstream	PLTP
NA	NA	NA	NA	NA	NA
NA	chr11:10486010-10486180	119	33170	Upstream	RNF141
NA	chr1:147481998-147482346	176	72631	Downstream	LOC388692
chr4:109760420-109760426	chr6:43403741-43403841	50	19233	Downstream	CRIP3
NA	chr15:41768914-41769414	501	2961	Downstream	CKMT1A
NA	NA	NA	NA	NA	NA
NA	chr9:34989141-34989291	72	9357	Upstream	DNAJB5
chr5:87993145-87993151	chr5:87992407-87992576	159	222204	Upstream	MEF2C
chr1:173259496-173259502	NA	NA	NA	NA	NA
chr3:47842646-47842652	chr3:47841127-47841433	299	46	Downstream	DHX30
NA	chr10:124899392-124899430	39	1765	Upstream	HMX2
NA	NA	NA	NA	NA	NA
NA	chr4:79321930-79322319	326	123811	Upstream	FRAS1
NA	NA	NA	NA	NA	NA

Second Best Hit

Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box	Best Location
340252	1311787	Downstream	ZNF588	51427	1	0	chr7:62451399-62451405
155185	102792	Upstream	TTYH3	80727	4	0.001	chr7:2740980-2740986
7091	1421580	Upstream	PSAT1	29968	3	1	chr9:81523617-81523623
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
170302	501160	Upstream	POLA1	5422	0	NA	NA
340252	1311787	Downstream	ZNF588	51427	1	0	chr7:62451399-62451405
152	288172	Upstream	OTOP1	133060	1	0	chr4:3991009-3991015
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
7767	19051	Downstream	ZNF225	7768	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
79230	50276	Upstream	MBD3L2	125997	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
170959	59743	Upstream	ZNF714	148206	0	NA	NA
6909	341855	Downstream	C17orf82	388407	1	0.008	chr17:56501577-56501583
89890	61627	Upstream	KBTBD7	84078	0	NA	NA
54834	0	Overlapping	WDR3	10885	3	0.001	chr1:118274393-118274399
79953	265131	Downstream	CST7	8530	0	NA	NA
81849	248032	Upstream	PIGK	10026	0	NA	NA
6013	34395	Downstream	RLN2	6019	0	NA	NA
389741	348783	Downstream	ANKRD20A1	84210	0	NA	NA
401262	29766	Upstream	SLC22A7	10864	0	NA	NA
21	64496	Downstream	RNPS1	10921	1	0.005	chr16:2322776-2322782
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
6595	606916	Downstream	VLDLR	7436	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
84105	43921	Downstream	CATSPER3	347732	0	NA	NA
147687	18843	Upstream	ZNF418	147686	2	0	chr19:63119804-63119810
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
79694	438404	Downstream	FUT9	10690	0	NA	NA
79937	66908	Downstream	LOC642265	642265	1	0	chr9:39278506-39278512
NA	NA	NA	NA	NA	NA	NA	NA
5725	23006	Upstream	PRG2	79948	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
81691	496	Downstream	EXOD1	112479	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
440157	239493	Upstream	POTE14	404785	1	0	chr14:18862824-18862830
2241	402687	Downstream	FBXL17	64839	0	NA	NA
57037	517	Downstream	BZW2	28969	0	NA	NA
81849	248032	Upstream	PIGK	10026	0	NA	NA
8535	40289	Downstream	CBX8	57332	1	1	chr17:75426285-75426291
NA	NA	NA	NA	NA	NA	NA	NA
51574	56735	Upstream	LARP7	51574	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
55003	28201	Downstream	TMEM14C	51522	1	0.012	chr6:10803190-10803196

Second Best Hit

Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box	Best Location
7562	67769	Downstream	ZNF493	284443	0	NA	NA
26683	54179	Upstream	OR4F16	81399	0	NA	NA
23332	206818	Upstream	MKI67IP	84365	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
9631	622	Downstream	NUP155	9631	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
3820	221346	Downstream	CLEC2D	29121	0	NA	NA
9064	8682	Upstream	FCN3	8547	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
51065	31648	Downstream	RAB8B	51762	0	NA	NA
51351	98757	Downstream	ZNF117	7670	1	0	chr7:64178750-64178756
NA	NA	NA	NA	NA	NA	NA	NA
576	55398	Upstream	SPOCD1	90853	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
84105	41805	Downstream	CATSPER3	347732	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
7673	23755	Upstream	ZNF230	7773	0	NA	NA
57653	38795	Downstream	ZNF322B	387328	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
389741	348792	Downstream	ANKRD20A1	84210	0	NA	NA
375513	138692	Upstream	GUSBL2	375513	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
7009	34271	Downstream	FMNL3	91010	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
2067	632	Downstream	ERCC1	2067	0	NA	NA
8073	83496	Downstream	KHDRBS1	10657	0	NA	NA
64779	11964	Downstream	FOXC2	2303	3	0	chr16:85146022-85146028
NA	NA	NA	NA	NA	NA	NA	NA
2690	343375	Upstream	SEPP1	6414	1	0	chr5:42504066-42504072
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
57212	4938	Downstream	CCDC27	148870	0	NA	NA
389741	348783	Downstream	ANKRD20A1	84210	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
28974	10309	Upstream	MGC3207	84245	1	0	chr19:13746443-13746449
9015	70677	Upstream	C1orf80	64853	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
84105	41807	Downstream	CATSPER3	347732	0	NA	NA
1539	889241	Downstream	SMC2	10592	0	NA	NA
56033	84721	Downstream	PTPDC1	138639	0	NA	NA
84172	42622	Downstream	CHCHD5	84269	0	NA	NA
5506	505265	Downstream	GPR85	54329	0	NA	NA
8622	202130	Upstream	WDR41	55255	2	0.539	chr5:76621717-76621723
NA	NA	NA	NA	NA	NA	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
340252	1311937	Downstream	ZNF588	51427	1	0	chr7:62451399-62451405
84839	8864	Upstream	MRPL54	116541	1	0.999	chr19:3722546-3722552
8573	422736	Upstream	GPR82	27197	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA

Second Best Hit

Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box	Best Location
80256	12839	Downstream	STOML2	30968	1	1	chr9:35106373-35106379
NA	NA	NA	NA	NA	NA	NA	NA
50862	51954	Upstream	AMPD3	272	0	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
8370	0	Overlapping	H4/o	554313	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
4437	3423	Upstream	DHFR	1719	1	1	chr5:79982702-79982708
9747	99714	Downstream	OR2F2	135948	1	0.001	chr7:143163369-143163375
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
55677	107817	Upstream	PROC	5624	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
28981	18043	Downstream	IFT81	28981	0	NA	NA
64147	508	Upstream	KIF9	64147	0	NA	NA
51002	35122	Downstream	NAT8B	51471	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
147687	18843	Upstream	ZNF418	147686	2	0	chr19:63119804-63119810
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
51065	31648	Downstream	RAB8B	51762	0	NA	NA
90317	49728	Downstream	PPP2R1A	5518	1	0	chr19:57334896-57334902
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
81849	248032	Upstream	PIGK	10026	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
284123	685682	Upstream	KCNJ12	3768	1	0	chr17:21944926-21944932
144193	328	Downstream	CCDC38	120935	1	0	chr12:94860961-94860967
94134	127872	Upstream	KIF5B	3799	2	0.837	chr10:32257803-32257809
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
6496	80881	Upstream	SIX2	10736	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
8726	57238	Downstream	C11orf73	51501	0	NA	NA
379	76282	Upstream	TMEM106A	113277	0	NA	NA
84105	43921	Downstream	CATSPER3	347732	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
3064	112048	Upstream	GRK4	2868	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
2067	1728	Downstream	ERCC1	2067	1	0.003	chr19:50620785-50620791
23	13698	Downstream	GNL1	2794	1	0	chr6:30647119-30647125
50862	51954	Upstream	AMPD3	272	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA

Second Best Hit

Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box	Best Location
284123	685682	Upstream	KCNJ12	3768	1	0	chr17:21944926-21944932
8341	323	Downstream	HIST1H2AK	8330	1	0.999	chr6:27914001-27914007
84105	40599	Downstream	CATSPER3	347732	0	NA	NA
26083	123456	Upstream	GOSR1	9527	0	NA	NA
4968	0	Overlapping	OGG1	4968	0	NA	NA
3659	52855	Upstream	IL5	3567	0	NA	NA
54331	128652	Downstream	C14orf166	51637	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
7767	18693	Downstream	ZNF225	7768	0	NA	NA
9811	437173	Downstream	ZBTB7C	201501	0	NA	NA
55672	129763	Downstream	LOC440570	440570	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
1131	705292	Downstream	FMN2	56776	1	0	chr1:237615922-237615928
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
57496	152179	Upstream	ERCC4	2072	0	NA	NA
79694	438404	Downstream	FUT9	10690	0	NA	NA
26683	54181	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
6638	214415	Upstream	SNURF	8926	2	0.418	chr15:22965758-22965764
NA	NA	NA	NA	NA	NA	NA	NA
7572	33487	Upstream	ZNF396	252884	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
83861	42133	Upstream	TAGAP	117289	0	NA	NA
84105	43921	Downstream	CATSPER3	347732	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
50862	51954	Upstream	AMPD3	272	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
51065	31648	Downstream	RAB8B	51762	0	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
221833	130875	Upstream	ABCB5	340273	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
7546	9250	Downstream	ZIC5	85416	0	NA	NA
114817	375125	Upstream	FLJ43276	388165	3	0.041	chr15:82540386-82540392
388567	20774	Upstream	VN1R1	57191	0	NA	NA
284123	689000	Upstream	KCNJ12	3768	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
6606	123	Upstream	SMN2	6607	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
6144	18607	Downstream	RASL11A	387496	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
79830	46900	Downstream	ZMYM6	9204	0	NA	NA
11128	4120	Downstream	RPS24	6229	1	0.266	chr10:79459309-79459315

Second Best Hit

Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box	Best Location
164022	1201626	Downstream	NBPF20	400818	0	NA	NA
80144	318999	Upstream	MRPL1	65008	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
4761	53307	Downstream	PPP1R1B	84152	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
6606	123	Upstream	SMN2	6607	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
161	59130	Downstream	CHID1	66005	1	0	chr11:960424-960430
26683	57173	Upstream	OR4F16	81399	1	0.082	chr1:553920-553926
NA	NA	NA	NA	NA	NA	NA	NA
1793	258275	Downstream	C10orf90	118611	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
84078	61342	Downstream	KBTBD6	89890	0	NA	NA
23167	576719	Upstream	KCNQ3	3786	1	0.006	chr8:132985062-132985068
NA	NA	NA	NA	NA	NA	NA	NA
80757	294398	Upstream	C14orf80	283643	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
57168	118549	Upstream	HPS4	89781	1	0.001	chr22:25086936-25086942
NA	NA	NA	NA	NA	NA	NA	NA
54969	170990	Upstream	MFAP3L	9848	0	NA	NA
5188	307148	Upstream	FBXW7	55294	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
9416	13398	Upstream	RND1	27289	0	NA	NA
641559	137075	Upstream	NBPF11	200030	1	0	chr1:144656682-144656688
375513	145419	Upstream	GUSBL2	375513	0	NA	NA
55672	129763	Downstream	LOC440570	440570	0	NA	NA
2596	510730	Downstream	ZBTB20	26137	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
960	222979	Upstream	PDHX	8050	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
170959	59743	Upstream	ZNF714	148206	0	NA	NA
10199	97	Downstream	MCEE	84693	1	0	chr2:71211434-71211440
NA	NA	NA	NA	NA	NA	NA	NA
6144	18607	Downstream	RASL11A	387496	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
284123	685682	Upstream	KCNJ12	3768	1	0	chr17:21944926-21944932
57168	118549	Upstream	HPS4	89781	1	0.001	chr22:25086936-25086942
26683	54181	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
440157	88540	Upstream	POTE14	404785	6	0	chr14:18711628-18711634
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
57709	235194	Upstream	CLDN11	5010	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
375387	71115	Downstream	WDR53	348793	1	0.021	chr3:197851563-197851569
55672	129763	Downstream	LOC440570	440570	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA

Second Best Hit

Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box	Best Location
55676	64050	Upstream	CARD12	58484	1	0.004	chr2:32279825-32279831
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
9274	21174	Downstream	CTF1	1489	1	0.266	chr16:30794542-30794548
26683	53390	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
641559	91551	Upstream	NBPF11	200030	1	0	chr1:144702045-144702051
57658	75267	Downstream	ATP5G2	517	0	NA	NA
26683	53397	Upstream	OR4F16	81399	0	NA	NA
51454	498152	Upstream	DIRC1	116093	1	0	chr2:188864671-188864677
10750	107530	Downstream	FAM83G	644815	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
3714	11755	Upstream	NUDT14	256281	0	NA	NA
476	152551	Upstream	CD58	965	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
221322	101264	Downstream	GJA1	2697	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
55149	67018	Downstream	MAP3K8	1326	0	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
54875	171578	Downstream	BNC2	54796	1	0	chr9:17032825-17032831
84105	43921	Downstream	CATSPER3	347732	0	NA	NA
170959	59740	Upstream	ZNF714	148206	0	NA	NA
284297	31531	Upstream	NAT14	57106	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
92140	131596	Downstream	LAPTM4B	55353	2	0.077	chr8:98725543-98725549
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
51326	0	Overlapping	LOC641522	641522	0	NA	NA
9158	1434	Downstream	CCDC85B	11007	0	NA	NA
637	3748	Downstream	BID	637	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
222	117981	Downstream	ALDH3B2	222	0	NA	NA
148738	108849	Upstream	NOTCH2NL	388677	0	NA	NA
10174	9814	Upstream	SORBS3	10174	0	NA	NA
164022	861796	Downstream	NBPF20	400818	0	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
343070	69316	Downstream	PRAMEF8	391002	0	NA	NA
5067	1443543	Downstream	ROBO2	6092	0	NA	NA
79818	35157	Downstream	ZNF587	84914	1	0	chr19:63017929-63017935
80256	12839	Downstream	STOML2	30968	1	1	chr9:35106373-35106379
26683	53390	Upstream	OR4F16	81399	0	NA	NA
136051	34703	Upstream	ZNF425	155054	0	NA	NA
26683	54181	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA

Second Best Hit

Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box	Best Location
NA	NA	NA	NA	NA	NA	NA	NA
57182	852557	Downstream	FAT4	79633	1	1	chr4:125604012-125604018
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
170959	59779	Upstream	ZNF714	148206	0	NA	NA
84105	43921	Downstream	CATSPER3	347732	0	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
353500	61355	Upstream	PABPC4	8761	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
1139	375484	Downstream	OTUD7A	161725	1	0.041	chr15:30109929-30109935
7703	4805	Downstream	PSMB3	5691	0	NA	NA
10867	252117	Upstream	TEAD4	7004	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
22932	116405	Upstream	UPK3B	80761	1	1	chr7:76094545-76094551
26683	53390	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
64750	141311	Upstream	LRRC37A3	374819	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
51506	5243	Downstream	USP21	27005	1	0	chr1:159390727-159390733
5433	26409	Upstream	MGC4268	83607	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
1482	85434	Upstream	BNIP1	662	0	NA	NA
55638	103589	Upstream	EBAG9	9166	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
84105	43915	Downstream	CATSPER3	347732	0	NA	NA
10743	71497	Downstream	PEMT	10400	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
836	34375	Downstream	CCDC111	201973	0	NA	NA
26683	57173	Upstream	OR4F16	81399	1	0.082	chr1:553920-553926
84105	43921	Downstream	CATSPER3	347732	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
10651	81299	Upstream	HOXD1	3231	0	NA	NA
431707	404006	Upstream	TYW3	127253	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
8665	23987	Downstream	NALP10	338322	1	0.063	chr11:7965392-7965398
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
2863	288737	Downstream	FLJ41821	401011	1	0	chr2:132743193-132743199
50509	4097	Downstream	RDH8	50700	1	0	chr19:9980610-9980616
NA	NA	NA	NA	NA	NA	NA	NA
3209	21660	Downstream	EVX1	2128	0	NA	NA

Second Best Hit

Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box	Best Location
NA	NA	NA	NA	NA	NA	NA	NA
81849	248032	Upstream	PIGK	10026	0	NA	NA
79830	71865	Upstream	SFPQ	6421	1	0	chr1:35358576-35358582
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
4437	3423	Upstream	DHFR	1719	1	1	chr5:79982702-79982708
26683	53390	Upstream	OR4F16	81399	0	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
5858	168096	Downstream	A2M	2	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
5360	19462	Upstream	PPGB	5476	0	NA	NA
26683	53392	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
51463	25425	Upstream	GJA8	2703	0	NA	NA
6007	19224	Downstream	C1orf63	57035	1	0.001	chr1:25465868-25465874
5324	989	Downstream	CHCHD7	79145	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
9953	60639	Downstream	CDRT15	146822	1	0	chr17:14141181-14141187
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
7582	616992	Downstream	BMS1L	9790	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
401152	21964	Upstream	FABP2	2169	0	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
23431	160546	Upstream	TNFAIP8L3	388121	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
26683	54178	Upstream	OR4F16	81399	0	NA	NA
83987	54668	Upstream	FLJ10781	55228	0	NA	NA
5605	41862	Downstream	ZBTB7A	51341	1	0	chr19:4059570-4059576
NA	NA	NA	NA	NA	NA	NA	NA
57085	44083	Upstream	C1orf187	374946	2	0.001	chr1:11718715-11718721
11273	1222	Upstream	ATXN2L	11273	0	NA	NA
164022	861796	Downstream	NBPF20	400818	0	NA	NA
284948	27152	Downstream	CAPG	822	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
134266	13035	Downstream	PCYOX1L	78991	1	0	chr5:148704707-148704713
84105	41807	Downstream	CATSPER3	347732	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
60529	215969	Upstream	EXT2	2132	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
54928	697191	Downstream	PENK	5179	1	0.925	chr8:58217978-58217984
26683	53390	Upstream	OR4F16	81399	0	NA	NA
6474	2200	Downstream	SHOX2	6474	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
79738	290358	Downstream	NAP1L1	4673	0	NA	NA
27089	1268	Downstream	GDF9	2661	1	0	chr5:132229470-132229476
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
7767	19051	Downstream	ZNF225	7768	0	NA	NA

Second Best Hit

Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box	Best Location
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
51621	201081	Upstream	OTUD7A	161725	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
60481	186217	Upstream	GCLC	2729	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
80256	12839	Downstream	STOML2	30968	1	1	chr9:35106373-35106379
57573	30485	Downstream	ZFP28	140612	1	0.998	chr19:61712022-61712028
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
28974	10309	Upstream	MGC3207	84245	1	0	chr19:13746443-13746449
4610	648	Upstream	MYCL1	4610	1	0.006	chr1:40139488-40139494
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
7091	1421580	Upstream	PSAT1	29968	3	1	chr9:81523617-81523623
NA	NA	NA	NA	NA	NA	NA	NA
50862	51954	Upstream	AMPD3	272	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
342096	30194	Downstream	ISLR2	57611	1	0.023	chr15:72179669-72179675
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
50862	51954	Upstream	AMPD3	272	0	NA	NA
3141	25584	Downstream	DSCR6	53820	0	NA	NA
4211	821714	Downstream	ETAA1	54465	0	NA	NA
3017	252	Upstream	HIST1H2BD	3017	2	0.001	chr6:26267131-26267137
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
79412	5503	Downstream	PAQR4	124222	2	0	chr16:2953823-2953829
NA	NA	NA	NA	NA	NA	NA	NA
79692	143747	Upstream	GUSBL1	387036	0	NA	NA
7673	23755	Upstream	ZNF230	7773	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
25897	6613	Downstream	RNF19	25897	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
55636	348019	Downstream	MGC34646	157807	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
50862	51954	Upstream	AMPD3	272	0	NA	NA
2690	343375	Upstream	SEPP1	6414	1	0	chr5:42504066-42504072
79692	143740	Upstream	GUSBL1	387036	0	NA	NA
1031	1261	Downstream	CDKN2C	1031	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
126961	534	Downstream	HIST2H3A	333932	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
26683	53390	Upstream	OR4F16	81399	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA

Second Best Hit

Gene1 ID	Gene2 Distance	Gene2 Direction	Gene2 Symbol	Gene2 ID	Total E-Boxes	Best E-Box	Best Location
8370	0	Overlapping	H4/o	554313	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
2619	200730	Downstream	FLJ45537	401535	1	0.003	chr9:88753119-88753125
128408	53504	Upstream	C20orf59	63910	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
5360	19454	Upstream	PPGB	5476	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
50862	51954	Upstream	AMPD3	272	0	NA	NA
388692	77361	Downstream	NBPF1	55672	0	NA	NA
401262	29766	Upstream	SLC22A7	10864	0	NA	NA
548596	40583	Downstream	CATSPER2	117155	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
25822	41810	Downstream	C9orf131	138724	0	NA	NA
4208	391986	Downstream	TMEM161B	153396	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
22907	21473	Upstream	DHX30	22907	1	0	chr3:47841157-47841163
3167	4429	Downstream	BUB3	9184	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
80144	319006	Upstream	MRPL1	65008	0	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA

Supplementary Table 2: Primer sequences for PCR reactions

Cyclin D2 forward	CCTTGACTCAAGGATGCGTTAGA
Cyclin D2 reverse	GAGCCGACTGCGGTGAAGT
Chr 22 E Box forward	TTACAGGTAAGCACCTCCATGACC
Chr 22 E Box reverse	GCAAAAGCTACCATTTAGGAACCC

Supplementary Table 3: Primer sequences for Q-PCR reactions

Cyclin D2 forward	CCTTGACTCAAGGATGCGTTAGA
Cyclin D2 reverse	GAGCCGACTGCGGTGAAGT
Chr 21EBox forward	AAAACGCAGCCACTTGGAGT
Chr 21EBox reverse	GCAGAAGTGCGGTTCATCAG
GDAP2 forward	GCAGAAATGCAAGAGAGAAAAGCT
GDAP2 reverse	CTGACTCCTCCTTGTTTGCTCA
RPL34 (1) forward	TCAAGAGTTGGGATTTAAACCCA
RPL34 (1) reverse	TTCGTTGTCATGGACACTGCA
RPL34 (2) forward	TACACACGTGCGCTTAATGTAAAATT
RPL34 (2) reverse	AACTACTTGAGGCTTTTCATGGGT
PHLPP froward	GCGGGATGGAACACTGAGA
PHLPP reverse	GGATTATGTGGCGATTCGGA
TCERG1 forward	CTTGCCCAGTGGAAATGCTC
TCERG1 reverse	TCTGTCGGCTTCCAGGAGAG
PKIB forward	GCATGTGGAACCCTTGCATAT
PKIB reverse	GCATTCCCTGCAGATGTGC
CDC14A (1) forward	CCTCGGTCAGGTCCTCCTCT
CDC14A (1) reverse	TTCTTGCAGAAACCTCATCCCT
CDC14A (2) forward	TCTGGGAGGGATGAGGTTTCT
CDC14A (2) reverse	TGGGTGGTAACATACGAAGGATC
OGG1 forward	ATGGCACTGGTCTGATGTTGATT
OGG1 reverse	TGCAGCGGATTGACCCTT
CDCA2 forward	GCTTACCCTGCAGCACAGAATTA
CDCA2 reverse	CACCCAGAGCAGCTCCTGAC
NOL6 forward	ACTCCAGGCCCAAGTCCAG
NOL6 reverse	GCGAGTAATCGACTGATAATCCAA
TRH forward	GCGCTTCTCTCAATCTCTCTGTC
TRH reverse	TCTGCACCACCTCTGGAGATCT
KLC4 forward	TTGCAGGTGAGTCTTTGAGGGTAT
KLC4 reverse	CATCTTTCTGGTAGTGCGAGAATAAG
GAS1 forward	GCAGCTCACATTCAGGCATTC
GAS1 reverse	TCCGAGCCTCCTTAAGAGTAGGA
PLTP forward	GCAGGTCCGGAATGGTGAT
PLTP reverse	GGTTTCTCTGCTGCTTCAGTGAA

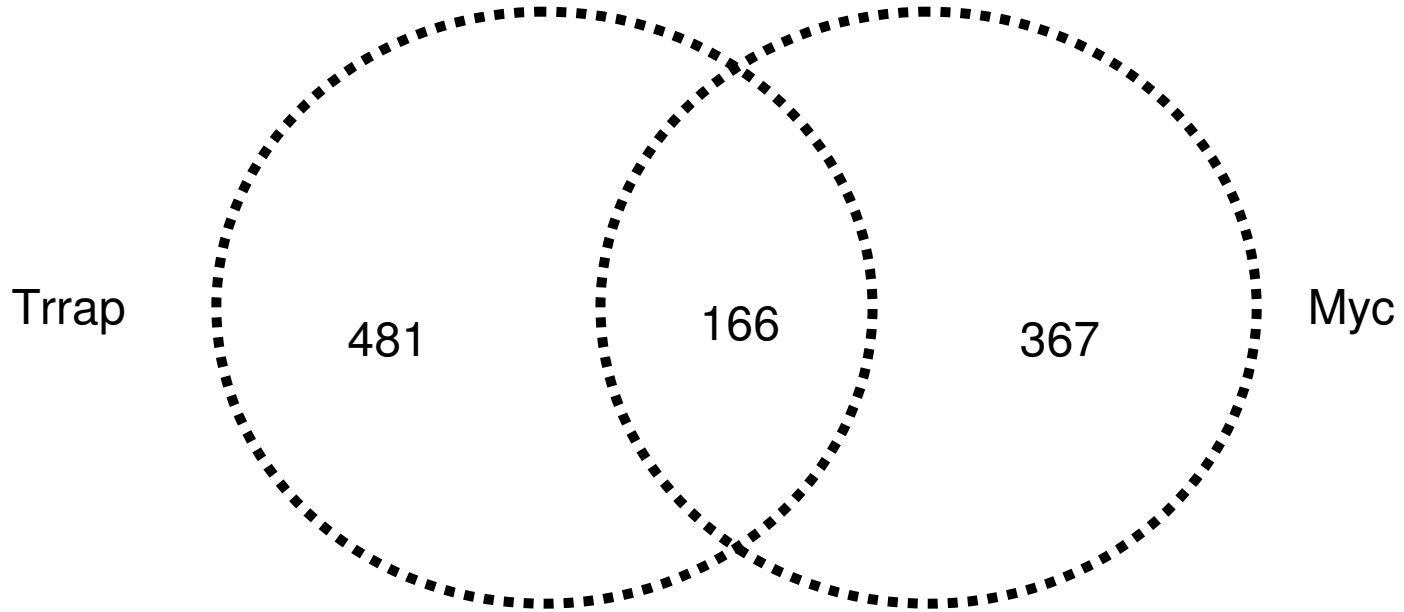


Figure 1A: Number of Overlapping Spots at P < 0.05

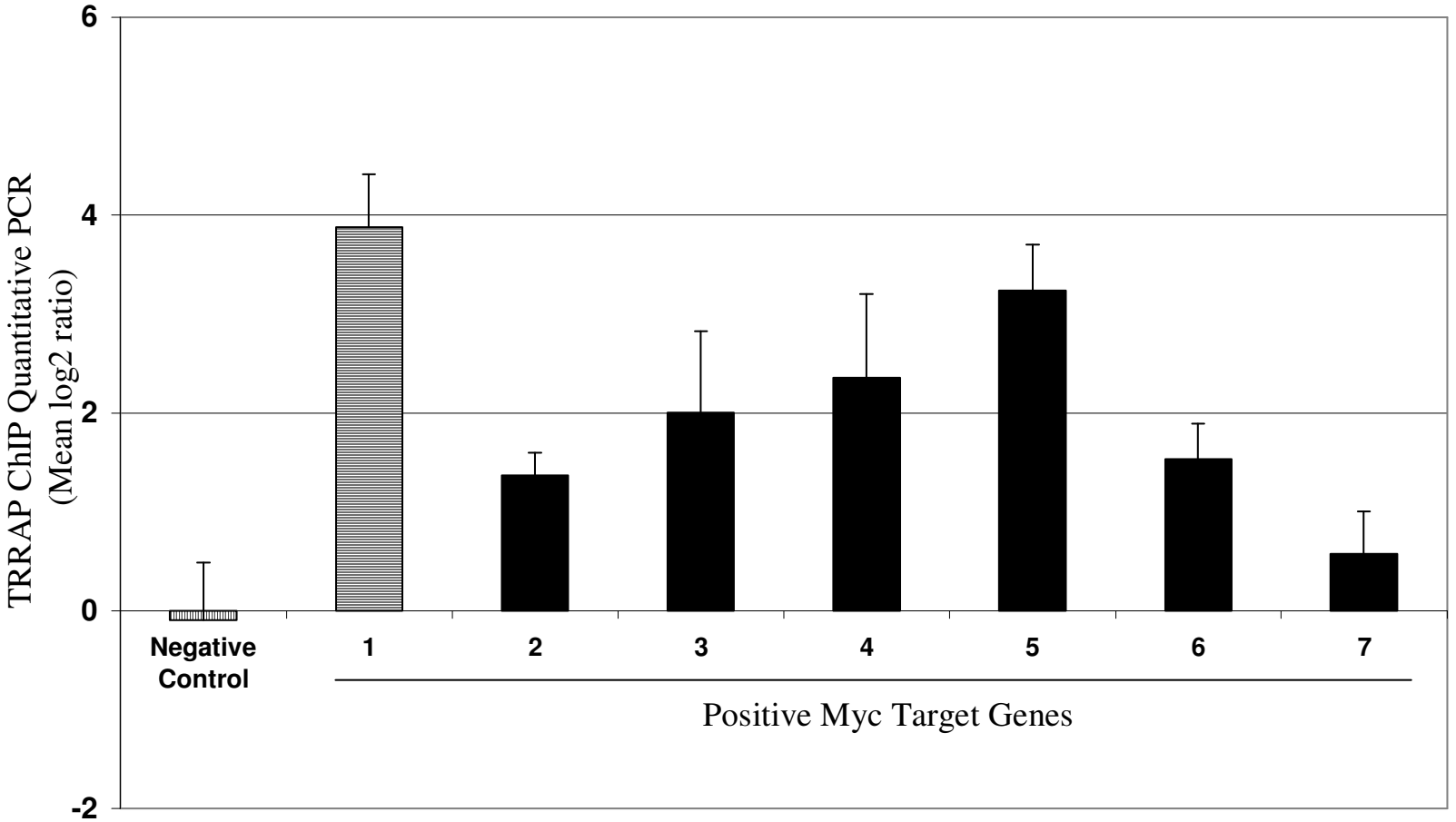


Figure 1B: TRRAP binds to Myc-specific target genes. Q-PCR analysis of TRRAP specific ChIP using PCR primers to established Myc negative and positive control target genes

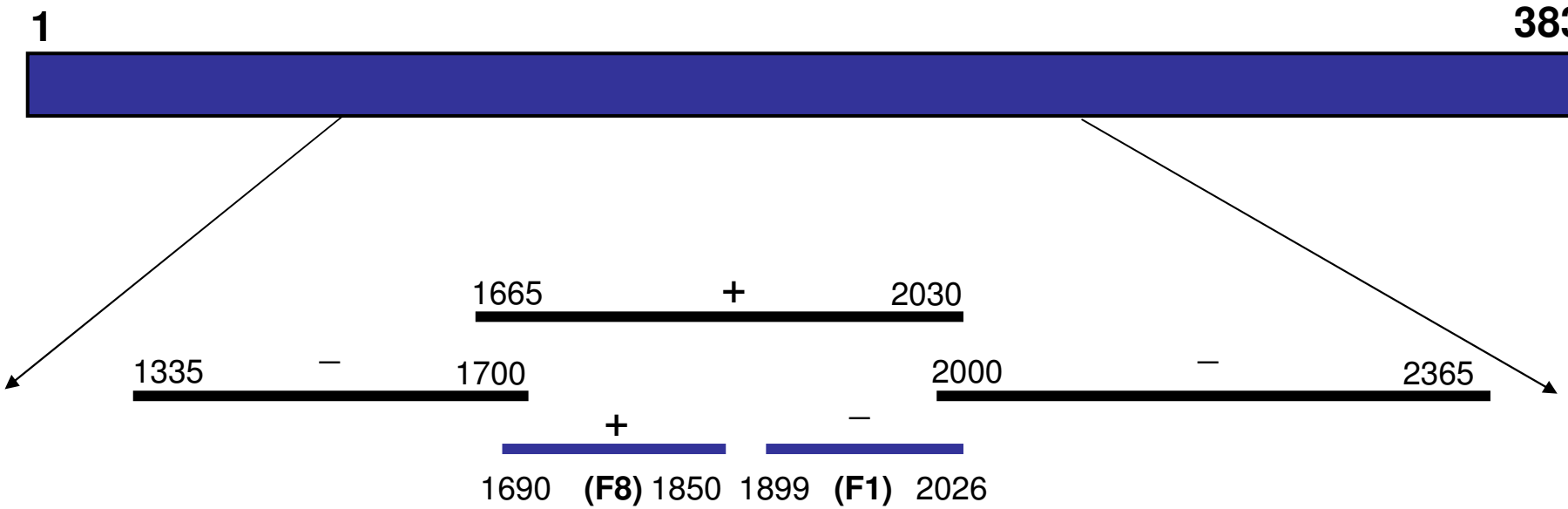


Figure 2A: Schematic diagram of TRRAP (3830 amino acids) highlighting the regions we have shown that do (+) and do not (-) interact with Myc using a variety of technologies to evaluate interaction both in vitro (e.g. GST-pulldown, circular dichroism) and in vivo (e.g. repressed transactivator assay, co-immunoprecipitation). From this analysis we identified Fragment F8 as a minimal Myc Binding Domain (MBD).

Figure: TRRAP Fragment MBD2 interacts with c-Myc in mammalian cells

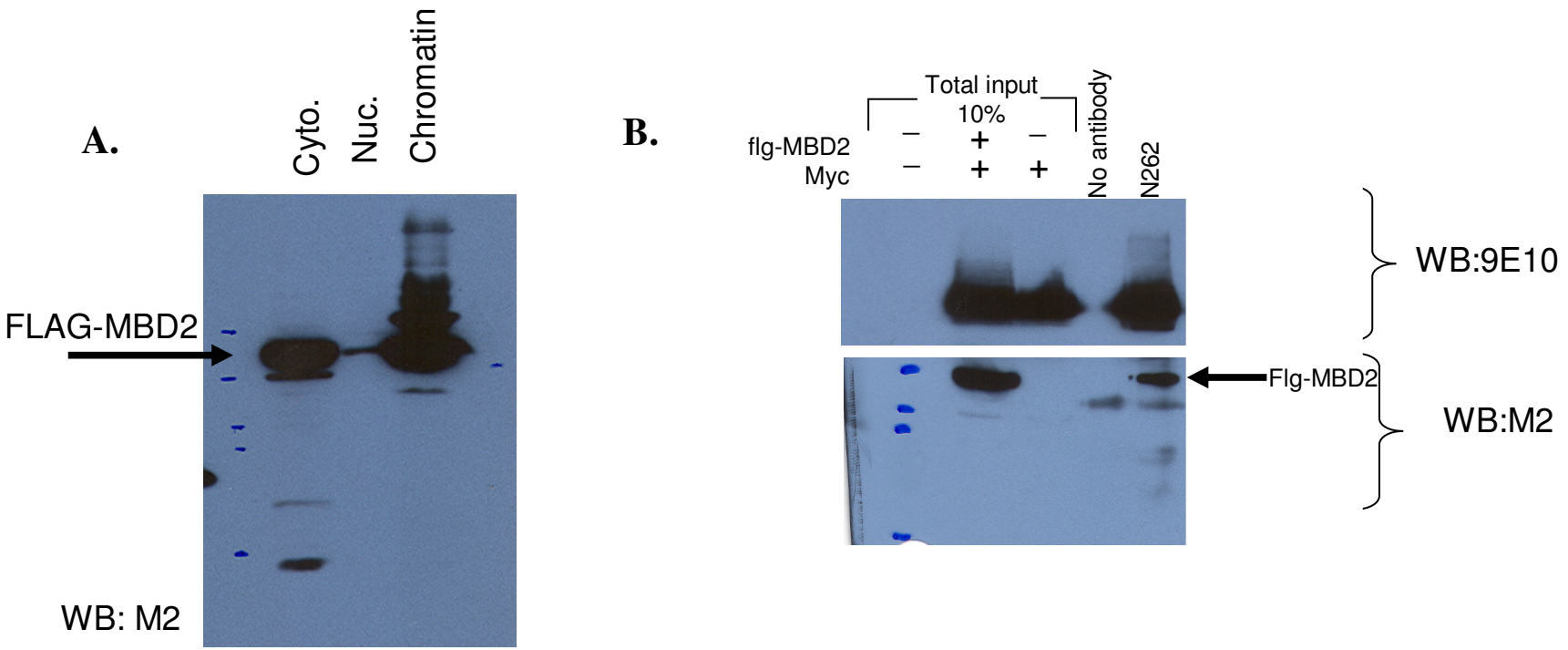


Figure 2B: Evidence that a second region of TRRAP has been identified as a Myc Binding Domain (MBD2). Ectopic expression of Flag-tagged MBD2 fragment in cells is evident in chromatin-enriched fraction (A) Immunoprecipitation of Myc using antibody directed to the N-terminal domain (N262) results in co-immunoprecipitation of both Myc, as detected by western blot (WB) using antibody 9E10, and Flag-MBD2, as detected by western blot using anti-Flag M2.